

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2002 年 3 月 21 日 (21.03.2002)

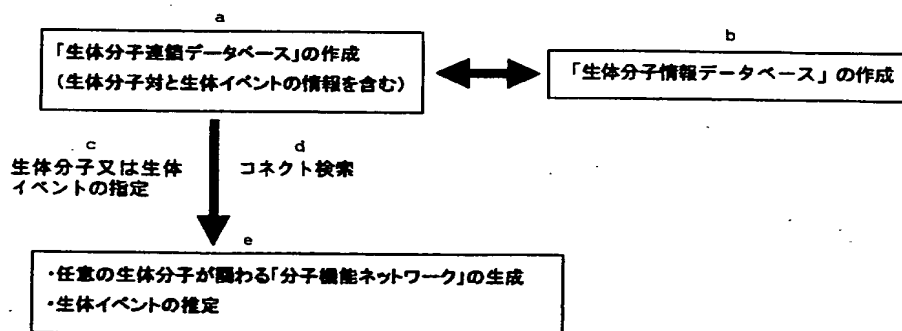
PCT

(10) 国際公開番号
WO 02/23395 A1

- (51) 国際特許分類: G06F 17/30, 17/60 (74) 代理人: 今村正純, 外(IMAMURA, Masazumi et al); 〒104-0031 東京都中央区京橋1丁目8番7号 京橋日殖ビル8階 Tokyo (JP).
- (21) 国際出願番号: PCT/JP01/07830
- (22) 国際出願日: 2001 年 9 月 10 日 (10.09.2001) (81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願2000-276699 2000 年 9 月 12 日 (12.09.2000) JP (84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (71) 出願人 (米国を除く全ての指定国について): 株式会社 医薬分子設計研究所 (INSTITUTE OF MEDICINAL MOLECULAR DESIGN, INC.) [JP/JP]; 〒113-0033 東京都文京区本郷5丁目24番5号 角川本郷ビル4F Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 板井昭子 (ITAI, Akiko) [JP/JP]; 〒113-0033 東京都文京区本郷5-16-6 Tokyo (JP).
- 添付公開書類:
— 国際調査報告書
- 2 文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(54) Title: METHOD OF FORMING MOLECULAR FUNCTION NETWORK

(54) 発明の名称: 分子機能ネットワークの生成方法



- a...FORMATION OF "BIOLOGICAL MOLECULAR CHAIN DATA BASE"
(INVOLVING BIOLOGICAL MOLECULAR BASE PAIR AND BIOLOGICAL EVENT DATA)
- b...FORMATION OF "BIOLOGICAL MOLECULAR DATA BASE"
- c...DESIGNATION OF BIOLOGICAL MOLECULE OR BIOLOGICAL EVENT
- d...CONNECT SEARCHING
- e...FORMATION OF "MOLECULAR FUNCTION NETWORK" IN WHICH ARBITRARY BIOLOGICAL MOLECULE PARTICIPATES ESTIMATION OF BIOLOGICAL EVENT

(57) Abstract: A method of forming a molecular function network involving biological events by the connect searching with the use of a biological molecular chain data base involving biological event data; and a method of estimating the pathway between an arbitrary biological molecule and an arbitrary biological event in the network, or estimating a biological event in which an arbitrary biological molecule in the network participates.

BEST AVAILABLE COPY

[続葉有]

BEST AVAILABLE COPY

WO 02/23395 A1



(57) 要約:

生体イベントの情報を含む生体分子連鎖データベースを用いて、コネクト検索を行なうことにより、生体イベントを含む分子機能ネットワークを生成する方法、及び該ネットワーク中の任意の生体分子と任意の生体イベントとの間の経路を推定し、又は該ネットワーク中の任意の生体分子が関わる生体イベントを推定する方法。

明 細 書

分子機能ネットワークの生成方法

技術分野

本発明は、生体イベントの情報を含む生体分子データベースの作成方法と利用方法に関する。

背景技術

生体中には、DNA、RNA、蛋白質、多糖類などの生体高分子の他に、アミノ酸、核酸、脂質、糖質、一般低分子化合物など多様な分子が存在して、それぞれの機能を担っている。生体系の特徴は、単に多様な生体分子から構成されているというだけでなく、機能の発現をはじめとする生体内のすべての現象が生体分子間の特異的な結合を通じて起きることにある。この特異的な結合では、共有結合は形成されずに、分子間力のみによって安定化された複合体が形成される。従って、生体分子はそれが単独で存在する状態と複合体状態の間で、平衡状態で存在することになるが、特定の生体分子間では複合体状態の安定化が大きいために、この平衡が複合体側に著しく偏っている。その結果、多数の他分子の存在下で、かなり希薄な濃度でも実質的に特定の相手を識別して結合することができる。酵素反応においては、基質は酵素との複合体を形成した状態で所定の化学変換を受けた後に反応生成物として放出されるし、シグナル伝達においては、メディエーター分子の標的分子への結合に伴う標的分子の構造変化を通じて細胞外のシグナルが細胞内に伝達される。

近年、ゲノム研究の進展は目覚ましく、ヒトをはじめとする多様な生物種についてゲノム配列が解明され、遺伝子や遺伝子産物である蛋白質の配列、臓器毎の蛋白質の発現、蛋白-蛋白相互作用等についてもゲノムワイドな体系的な研究が進みつつある。それらの研究成果の大部分はデータベースとして公開されて、世

界中の利用に供されている。遺伝子や蛋白質の機能、疾患の原因又は背景となる遺伝子の推定、遺伝子多型との関わりについても少しずつ解明が進み、遺伝情報に基づく医療や新薬開発への期待が高まってきた。

一方、遺伝情報の担い手は核酸であるが、エネルギー代謝、物質変換、シグナル伝達など生命機能の殆どは、核酸以外の分子が担っている。蛋白質は遺伝子という設計図に基づいて直接生産される点で他のカテゴリーの分子と異なり、その種類も多い。酵素も、低分子性の生体内生理活性化合物の標的生体分子も、蛋白質の生体内生理活性化合物の標的生体分子も(糖で修飾されていることも多いが)、すべて蛋白質である。疾患の根本原因はともかくとして、多くの疾患や症状が蛋白質や低分子化合物の量やバランス、場合によっては質(機能)の異常の結果であると考えられている。既存医薬の殆どが、蛋白質を標的としてその機能を制御する化合物である。蛋白質と違い、核酸の立体構造が低分子医薬の標的として特異性を発揮しにくいという理由もあるが、抗生物質や抗菌剤、或いは農薬の殺虫剤や抗カビ剤に至るまで、標的は蛋白質である。

従って、遺伝情報に基づいた医療や新薬開発を行うには、生体内のそれぞれの蛋白質と低分子化合物の機能とそれらの分子間の特異的関係を明らかにする必要がある。さらに、異なる酵素が次々に関与して必要な分子が生合成されることや、異なる分子が次々に結合することでシグナルが伝達されることなどから、これらの分子は機能上や生合成上で互いに直接又は間接のつながりをもっており、そのつながり(分子機能ネットワーク)の情報が重要である。さらに、これまでの研究で、メディエーターやホルモンなど多彩な臨床症状や生理現象や生体反応の発現に直接関わっている分子が多数明らかにされており、適切な治療には分子機能ネットワークとの関係付けが不可欠である。また、創薬戦略において、副作用リスクを考慮しつつ適切な新薬開発の標的分子を設定するためには、標的分子を含む分子機能ネットワークの考慮が必要である。

蛋白質に関連したデータベースとしては、SwissProt (the Swiss Institute of Bioinformatics (SIB), European Bioinformatics Institute (EBI))、PIR

(National Biomedical Research foundation (NBRF))があり、いずれも配列情報の他、生物種、機能、機能メカニズム、発見者、文献その他のアノテーションの情報をもっている。

分子のつながりに注目した分子ネットワーク型データベースのうち、代謝経路に関するものとしては、KEGG (金久ら、京都大学)、Biochemical Pathways (Boehringer Mannheim), WIT (Russian Academy of Sciences), Biofrontier (呉羽化学)、Protein Pathway (AxCell), bioSCOUT (LION), EcoCyc (DoubleTwist), UM-BBD (Minnesota Univ.) がある。

KEGGのPATHWAYデータベースにはメタボリックパスウェイとシグナルトランスダクションパスウェイがあり、前者は物質代謝、エネルギー代謝に関わる一般低分子化合物の代謝経路を、後者はシグナル伝達系の蛋白質を扱っている。いずれも予め定義された分子ネットワークを静的なGifファイルで提供している。前者では、酵素とリガンドの情報を別のテキスト形式の分子データベースLIGAND (金久ら、京都大学) 及び ENZYME (IUPAC-IUBMB) から取り込むようになっている。生理活性ペプチドの生成に関わる酵素や標的生体分子は含まれない。

EcoCycは大腸菌の物質代謝をデータベース化したもので、個々の酵素反応に関するデータと、既知のパスウェイに関するデータ (該パスウェイに属する酵素反応の集合として表される) に基づいて、パスウェイの図式的な表示を行なっている。EcoCycの検索機能としては、分子名やパスウェイ名に対する文字列や略号による検索法が提供されているが、任意の分子を指定して新たなパスウェイを検索することはできない。

シグナル伝達に関するものとしては、CSNDB (国立医薬品食品衛生研究所、日本)、SPAD (久原ら、九州大学)、Gene Net (Institute of Cytology & Genetics Novosibirsk, Russia), GeNet (Maria G. Samsonova) がある。

蛋白-蛋白相互作用のデータベースとしては、DIP (UCLA), PathCalling (CuraGen), ProNet (Myriad) がある。

遺伝子・蛋白発現のデータベースとしては、BodyMap (東京大学、大阪大学)、

SWISS-2DPAGE (Swiss Institute of Bioinformatics), Human and mouse 2D PAGE database (Danish Centre for Human Genome Research), HEART-2DPAGE (GermanHeart)、PDD Protein Disease Databases (NIMH-NCI), Washington University Inner Ear Protein Database (Washington Univ.), PMMA-2DPAGE (Purkyne Military Medical Academy), Mito-Pick (CEA, France), Molecular Anatomy Laboratory (Indiana University), Human Colon Carcinoma Protein Database (Ludwig Institute for Cancer Research) がある。

生体反応シミュレーション型の分子ネットワークとしては E-Cell (冨田ら、慶応大学), e E.coli (B. Palsson), Cell (D. Lauffenburger, MIT), Virtual Cell (L. Leow, Connecticut Univ.), Virtual Patient (Entelos, Inc.) がある。

生体分子と機能の関係については、広範な蛋白の情報を集めた SwissProt の他、COPE (University of Munich) がサイトカインの機能情報をテキスト形式で公開している。ARIS ((株) 日本電子計算) は、医学・薬学分野を中心に約 400 の国内雑誌、20 の海外雑誌から医薬品の副作用、相互作用、農薬・化学物質による中毒に関わる文献情報を収録しているが、生体分子の生理的作用や細胞レベル以上での応答についてのデータベースはこれまでにない。遺伝子と疾患については、遺伝病と蛋白質のアミノ酸変異の情報を集めた OMIM (NIH) がある。いずれもテキスト形式のデータが記述されていて、キーワードで検索できる。

分子間のつながりに注目した既存データベースの問題点は以下の点にある。分子ネットワーク型データベースは含まれる分子と分子間のつながりが分かっている系に対して作られており、予め分子間の関係を考慮して配置することが可能なために、静的な Gif 等での表現で済んできた。しかし、この方式では、新しい分子や分子間のつながりの追加が困難である。今後存在が明らかにされるであろう分子も含めると 10 万以上あり (KEGG で扱っている分子数は医薬分子を含めて約 1 万)、さらにそれらの分子間の関わりが今後の研究により解明されると、分子ネットワークの複雑さは加速度的に増すことが予想される。新しい分子の追加に対応でき、膨大な分子と分子間の関係の情報を保持しながら、必要な情報を含む部

分的分子ネットワークを生成できる新しい方法が必要である。

2001年9月7日現在、KEGGでは分子間のつながりを2分子の対の情報として保存し、この情報を用いてメタボリックパスウェイ中の任意の2分子間をつなぐ経路を探索することが可能になっている。しかし、このような経路探索問題には、2分子間をつなぐ経路が長くなるにつれて計算時間が級数的に増加するという問題がある。

一方、テキスト形式のデータベースでは分子データの追加に限界はない。しかし、個々の分子のデータから機能上又は生合成上関連ある分子を次々に探索を繰り返して、多数の分子のつながりを示す分子ネットワークを生成するには困難が伴う。検索時に必要な分子についてのつながりが動的にかつ自動的に得られるようなデータの保持方法と検索方法を開発する必要がある。また、疾患や病態を分子レベルで理解するためには、生体分子やそのネットワークと生体の応答や生理作用との関係を記述する新しい工夫が必要である。

発明の開示

本発明の課題は、多様な生体応答や現象を生体分子の機能とその分子間の関わりにおいて理解するための仕組みと方法を提供することであり、より具体的には生体分子の情報を生体応答と結びつけることのできるデータベースとその検索法を提供することにある。さらには、それらの膨大な情報から、任意の生体応答又は生体分子に関連するシグナル伝達経路や生合成経路のみを迅速に効率よく取り出し、有望な創薬ターゲットや副作用リスクを推定するための方法を提供することも本発明の課題の1つである。

本発明者らは、上記の課題を解決すべく鋭意努力した結果、直接結合する生体分子の対をパーツとした情報を集積することで生体分子間のつながりを網羅し、生体分子の中で生体応答の発現に直接関わるキイ分子については標的生体分子との対の情報に生理作用、生体応答、臨床症状等からなる生体イベントの情報を加えて保存し、指定した1以上の任意の生体分子又は生体イベントを含むつながり

を自動的に次々と探索して分子機能ネットワークを作成することによって上記の課題を解決できることを見出した。

すなわち、本発明により、直接結合する生体分子対に関する情報を蓄積した生体分子連鎖データベースを用いて分子機能ネットワークを生成する方法が提供される。この発明の好ましい態様によれば、生体イベントの情報をもつ生体分子連鎖データベースを用いて生体イベントの情報と関係付けた分子機能ネットワークを生成する上記の方法；生体分子自身に関する情報を含む生体分子情報データベースを用いる上記の方法；及び生体イベントの情報と関係付けた医薬分子を含む分子機能ネットワークを生成する上記の方法が提供される。また、直接結合する生体分子対に関する生体イベントの情報を蓄積した生体分子連鎖データベースを用いて任意の生体分子又は医薬分子が直接又は間接に関わる生体イベントを推定する方法も本発明により提供される。さらに、生体分子が、外部のデータベース又は文献中の遺伝子でコードされる蛋白質である場合に、該生体分子の分子略号と該遺伝子の名称又は ID 又は略称を対応付けるデータベースを作成することにより、遺伝子の多型や発現の情報を分子機能ネットワークを用いて解析する方法が本発明により提供される。

本発明のさらに好ましい態様によれば、ネットワーク上での連なりに基づいてグループ化した生体分子対データをサブネットとして、帰属するサブネット及びサブネット間の包含関係に基づいて、分子機能ネットワークを階層化することを特徴とする上記の方法；生体分子対に関する情報を該分子対の帰属経路名・帰属サブネット名等に基づいて階層化して保存することを特徴とする上記の方法；生体分子自身に関する情報を遺伝子群からの発現パターンや細胞表面への出現パターン等に基づいて階層化して保存することを特徴とする上記の方法；生体イベントに関する情報を該イベントの上位概念による分類や病態イベントとの関連に基づいて階層化して保存することを特徴とする上記の方法も提供される。さらに、生体分子対に関する上位階層、生体分子自身に関する上位階層、生体イベントに関する上位階層のそれぞれの保存項目に対して、項目間の関連性や依存関係等に

関する情報を保存することを特徴とする上記の方法；生体分子情報データベース又は生体分子連鎖データベースに保存された階層化の情報を利用することにより、分子機能ネットワークの生成を容易にすることを特徴とする上記の方法；生体分子情報データベース又は生体分子連鎖データベースに保存された階層化の情報を利用することにより分子機能ネットワークの表現での詳細さ制御することを特徴とする上記の方法も、本発明により提供される。

さらに、本発明により、以下に示す方法及びデータベースが提供される。

1. 生体イベントの情報と生体分子を関係付ける方法。
2. 生体イベントの情報と関係付けた分子機能ネットワークの生成方法。
3. 生体イベントの情報と関係付けた医薬分子を含む分子機能ネットワークの生成方法。
4. 任意の生体分子が直接又は間接に関わる生体イベントを推定する方法。
5. 生体イベントの情報をもつ生体分子連鎖データベースを用いて、任意の生体分子が直接又は間接に関わる生体イベントを推定する方法。
6. 生体イベントの情報をもつ生体分子連鎖データベースを用いて、任意の生体分子に関わる分子機能ネットワーク及び該分子が直接又は間接に関わる生体イベントを推定する方法。
7. 直接結合する生体分子対の情報に、生体イベントの発現に直接関わるキイ分子とその標的分子の対及び該生体イベントの情報を加えた生体分子連鎖データベース。
8. キイ分子に由来する生体イベントの情報を含む生体分子連鎖データベース。
9. 生体イベントの情報をもつキイ分子を含む生体分子連鎖データベース。
10. 生体分子連鎖データベースのコネクト検索で得られる分子機能ネットワーク。
11. 上記7から9の生体分子連鎖データベースのいずれか一つを用いて、任意の生体分子に関わる分子機能ネットワークと生体イベントを推定する方法。
12. 上記7から9の生体分子連鎖データベースのいずれか一つと医薬分子連鎖

データベースを用いて、任意の生体分子又は医薬分子が関わる分子機能ネットワークと生体イベントを推定する方法。

13. 生体イベントの情報にキイ分子の量的又は質的な変動に対応したアップオアダウンの情報を含む、上記1から12に記載の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

14. 生体イベントの情報にキイ分子の生成臓器と生体イベントの発現臓器の情報を含む、上記1から12に記載の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

15. 生体イベントの情報にキイ分子の量的又は質的な変動に対応したアップオアダウンの情報ならびにキイ分子の生成臓器と生体イベントの発現臓器の情報を含む、上記1から12に記載の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

16. 直接結合する生体分子対と該結合の関係を表わす情報を保存することにより、1以上の任意の生体分子が機能上又は生合成上、直接又は間接に関係する分子機能ネットワークを作成する方法。

17. 直接結合する生体分子対の情報の集合を用いて、任意の生体分子と機能上又は生合成上、直接又は間接に関わるキイ分子を探す方法。

18. 請求項17に記載の方法に基づき、任意の生体分子が直接又は間接に関わる生体イベントを推定する方法。

19. 直接結合する生体分子対と該結合の関係を表わす情報を保存することにより、生体分子間の機能上又は生合成上の関わりを示す分子機能ネットワークを作成する方法。

20. 直接結合する生体分子対と該結合の関係を表わす情報をパーツとして保存し、コネクト検索することにより、任意の1以上の生体分子に関わる分子機能ネットワークを作成する方法。

21. 直接結合する生体分子対と該結合の関係を表わす情報をパーツとして保存し、コネクト検索することにより、指定する1以上の生体分子に生合成上又は機

能上、直接又は間接に関連する生体分子群を抽出する方法。

22. 疾患に関連する生体イベント群に基づき、該疾患が関連する分子機能ネットワークを推定する方法。

23. 疾患に関連する生体イベント群に基づき、該疾患が関連する分子機能ネットワークを推定し、可能な創薬ターゲットを推定する方法。

24. 疾患に関連する生体イベント群に基づき、該疾患が関連する分子機能ネットワーク上の生体分子を創薬ターゲットとした場合の副作用リスクを推定する方法。

25. ある疾患に関連した分子機能ネットワーク上の任意の生体分子の機能の制御による生体イベントのアップオアダウンを予測する方法。

26. キー分子の量的変動と生体イベントのアップオアダウンの情報を利用して、創薬ターゲットの選定を支援する方法。

27. 上記26に記載の方法で用いる生体分子連鎖データベース。

28. 医薬分子と標的生体分子の対の情報を含む生体分子連鎖データベース。

29. 医薬分子と標的生体分子の対及び作用と副作用の情報を付加した生体分子連鎖データベース。

30. 医薬分子と標的生体分子の対及び作用と副作用の情報を付加した生体分子連鎖データベースを用いて、医薬化合物の副作用リスク又は薬剤間相互作用を推定又は回避する方法。

31. 医薬分子と標的生体分子の対及び作用と副作用の情報を付加した生体分子連鎖データベースを用いて、必要に応じて遺伝子多型の情報とリンクして、疾病治療のための医薬化合物の選択及び用量設定を行う方法。

32. 生体分子連鎖データベース又は分子機能ネットワーク中の蛋白質が遺伝子データベースとリンクされていることを特徴とする、上記1から31に記載の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

33. 生体分子連鎖データベース又は分子機能ネットワークが、ゲノム配列に対応付けられた遺伝子の情報とリンクされていることを特徴とする、上記1から3

1の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

34. 生体分子連鎖データベース又は分子機能ネットワークが、臓器毎の蛋白発現情報と対応付けられた遺伝子の情報とリンクされていることを特徴とする、上記1から31の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

35. 生体分子連鎖データベース又は分子機能ネットワークが、遺伝子多型に関わる遺伝子の情報とリンクされていることを特徴とする、上記1から31の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

36. 生体分子連鎖データベース又は分子機能ネットワークが、他生物種のゲノム又は遺伝子配列と対応が付けられたゲノム又は遺伝子の情報とリンクされていることを特徴とする、上記1から31の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

37. 医薬分子の投与による特定の臓器における蛋白発現の変動の情報を用いて疾患メカニズムを推定するための、上記1から31の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

38. 特定疾患に頻度高く見られる遺伝子多型群の情報を解析するために用いる上記1から31の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

39. 生体分子対の関係がパターン化されたものであることを特徴とする、上記16から21の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

40. 生体イベントがパターン化されて表現されることを特徴とする、上記1から31の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

41. キー分子の量的変動に対する生体イベントのアップオアダウンの情報がパターン化されることを特徴とする、上記13から15の方法又は生体分子連鎖データベース又は分子機能ネットワーク。

42. 必要に応じて2以上の生体分子を1つの仮想生体分子として扱うことを特徴とする、上記1から41に記載の方法又は分子連鎖データベース又は分子機能ネットワーク。

43. 分散して配置された1以上の生体分子連鎖データベースを通信を介して利

用することを特徴とする、上記 1 から 4 1 に記載の方法又は分子連鎖データベース又は分子機能ネットワーク。

4 4. 生体イベントの発現に直接関わる生体分子の情報を含むデータベースを作成し、生体イベントの情報を必ずしも含まない分子機能ネットワークのデータベースと併用することを特徴とする、上記 1 から 4 1 に記載の方法又は分子連鎖データベース又は分子機能ネットワーク。

4 5. 生体イベントの情報を必ずしも含まない分子機能ネットワークのデータベースから、任意の分子に関わる部分的な分子機能ネットワークを抽出し、該ネットワークを構成する分子に基づいて、生体イベントの発現に直接関わる生体分子の方法を含むデータベースを検索することを特徴とする、上記 1 から 4 1 に記載の方法又は分子連鎖データベース又は分子機能ネットワーク。

4 6. 生成臓器又は作用臓器その他の情報により対象とする生体分子又は生体分子対を絞り込んだ生体分子連鎖データベース、又はそのデータベースを用いて作成した分子機能ネットワーク又は分子機能ネットワークを生成する方法

4 7. 生体分子機能データベースをコネクト検索していったん生成した分子機能ネットワークを、各ネットワークに含まれる生体分子や生体イベントの情報その他により、さらに絞り込む方法又は絞り込んで作成した分子機能ネットワーク

4 8. 生成臓器又は作用臓器その他の情報により対象とする生体分子又は生体分子対を絞り込んだ生体分子連鎖データベースを用いて、作成した分子機能ネットワークを、各ネットワークに含まれる生体分子や生体イベントの情報その他により、さらに絞り込む方法又は絞り込んで作成した分子機能ネットワーク

4 9. 上記 1 から 4 8 に記載の方法を実施するためのプログラムとデータベースからなるコンピュータシステム。

5 0. 上記 1 から 4 8 に記載のデータベースを記録したコンピュータ読み取り可能な媒体。

5 1. 上記 1 から 4 8 に記載の分子機能ネットワークに関する情報を記録したコンピュータ読み取り可能な媒体。

52. 上記1から48に記載のデータベースと、上記1から48に記載の方法を実施するためのプログラムを記録したコンピュータ読み取り可能な媒体。

53. 階層化された生体イベントの情報と生体分子を関連付ける方法

54. 階層化された生体イベントの情報と関連付けた分子機能ネットワークの生成方法。

55. 生体分子対の情報を階層化して保存することを特徴とする分子機能ネットワークの生成方法。

56. 生体分子の集合状態を階層化して保存することを特徴とする分子機能ネットワークの生成方法。

57. 階層化して保存された生体分子対の情報に対して生体イベントを関連付ける方法。

58. 階層化して保存された生体分子の集合状態の情報に対して生体イベントを関連付ける方法。

59. 遺伝子群の転写の情報を階層化して保存することを特徴とする分子機能ネットワークの生成方法。

60. 蛋白発現の情報を階層化して保存することを特徴とする分子機能ネットワークの生成方法。

61. データベース中の任意のデータ項目に対してキーワード・数値パラメタ・分子構造・アミノ酸配列・塩基配列などに基づく検索を行ない、該検索結果に基づいて分子機能ネットワークを生成する方法。

62. 作成した分子機能ネットワークについて、そこに含まれる生体分子、生体分子対、生体イベントのデータに対して、キーワード・数値パラメタ・分子構造・アミノ酸配列・塩基配列などに基づく検索を行なうことにより、該ネットワークの部分集合を得る方法。

63. 作成した分子機能ネットワークについて、そこに含まれる生体分子、生体分子対、生体イベントのデータに対して、キーワード・数値パラメタ・分子構造・アミノ酸配列・塩基配列などに基づく検索を行なうことにより、該生体分子、生

体分子対、生体イベントを強調して表示する方法。

図面の簡単な説明

第1図は、本発明の方法の基本的な概念を表わす図である。

第2図は、本発明の方法で医薬分子連鎖データベースを用いる場合の概念を表わす図である。

第3図は、本発明の方法で遺伝情報データベースを用いる場合の概念を表わす図である。

第4図は、実施例1で対象としたレニン-アンジオテンシン系の概念を示す図である。

第5図は、実施例1の生体分子情報データベースの内容を示す図である。

第6図は、実施例1の生体分子連鎖データベースの内容を示す図である。

第7図は、実施例1の生体分子に関する検索で得られた分子機能ネットワークを示す図である。質問に用いた生体分子と生体イベントを太枠で示す。

第8図は、実施例1の医薬分子情報データベースの内容を示す図である。

第9図は、実施例1の医薬分子連鎖データベースの内容を示す図である。

第10図は、実施例1の医薬分子に関する検索で得られた分子機能ネットワークを示す図である。質問に用いた医薬分子と生体イベントを太枠で示す。

第11図は、実施例2の分子機能ネットワークの検索・表示プログラムの流れ図である。

第12図は、実施例2のコネクト検索(1点指定)の入力項目を示す図である。

第13図は、実施例2のコネクト検索(2点指定)の入力項目を示す図である。

発明を実施するための最良の形態

本明細書における用語の意味又は定義は以下の通りである。

「生体」とは、例えばオルガネラ、細胞、組織、臓器、生物個体、又は集合体など、生物に寄生する生命体を含む概念とする。

「生体イベント」とは、生体において内因的に又は外因的に現れるすべての現象、応答、反応、症状を含む概念とする。具体的な例として、転写、細胞の遊走、細胞の接着、細胞分裂、神経回路興奮、血管収縮、血圧上昇、血糖低下、発熱、痙攣、異種生物及びウイルスなど寄生体による感染その他を挙げることができる。また、光や熱などの生体外部からの物理的な刺激に対する応答も生体イベントの概念に含めることができる。

「病態イベント」とは「生体イベント」に含められる概念であって、「生体イベント」が量的又は質的にある閾値を超え、疾患又は病態と判断できる状態をいう。例として、血圧上昇の「生体イベント」が異常に亢進した結果の「病態イベント」としては高血圧又は高血圧症、血糖が正常範囲に制御できなくなった「病態イベント」としては高血糖又は糖尿病を挙げることができる。また、上記の例のように単一の生体イベントに関連するものだけではなく、複数の種類の生体イベントが関連している病態イベントもある。

「生体分子」とは、生体中に存在する核酸、蛋白質、脂質、糖質、一般低分子化合物その他のあらゆる構造の有機分子及びその集合体を指し、金属イオン、水、プロトンを含んでいてもよい。

「キイ分子」とは、生体分子のうち、主としてメディエーター、ホルモン、神経伝達物質、オータコイド等の分子群を指す。多くの場合、体内に特定の標的生体分子が存在して、その分子への直接結合が上記の「生体イベント」の引き金となることが知られている。これらの分子は生体内で生成されて作用を発現しているが、一般には生体系の外部から与えた場合にもその量に対応した生体イベントを発現する。具体的例として、アドレナリン、アンジオテンシン II、インシュリン、エストロゲンその他を挙げることができる。

「標的生体分子」とは、メディエーター、ホルモン、神経伝達物質、オータコイド等の生体分子又は医薬分子の受け皿となる特定の生体分子をいう。直接結合して特定のイベントを発現する。

「生体イベントのアップオアダウンの情報」とは、キイ分子又は標的生体分子

の量的又は質的な変動に対応して生体イベントが亢進・上昇又は低下・減少する等の情報をいう。キイ分子の量が一定の閾値を超えた場合にはじめてその生体イベントが起きる場合も含む。

「分子略号」とは、分子名称の代わりに分子を識別又は指定する目的に付するもので、各分子にユニークに対応している必要がある。分子名を短縮した略号でも、分子名と無関係な英数字列でもよいが、短い文字列であることが望ましい。既に世界的に使用されている分子略号がある分子の場合には、それを用いるのが望ましい。1ヶの分子に対して異なる方式で付けられた複数の分子略号を与え、構造グループや機能などにより階層化することも可能である。

「直接結合」とは、共有結合によらない分子間力によって安定な複合体を形成する、又は形成可能であることをいう。稀に共有結合が形成されることもあるが、この場合を含めた概念である。「相互作用」と呼ばれることも多いが、相互作用の方が広い意味を含んでいる。

「生体分子対」とは、生体中で直接結合できるか直接結合することが推定される生体分子の対をいう。具体的な例として、エストラジオールとエストロゲン受容体、アンジオテンシン変換酵素とアンジオテンシンIなどを挙げることができる。酵素反応における酵素と生成物の分子対の場合には、その複合体はあまり安定とは言えないが、生体分子対に含めるものとする。また、例えば two-hybrid 実験法で相互作用があるとされた2ヶの蛋白分子のように両者の相互の役割が明確でない場合を含めてもよい。光、音、温度変化、磁気、重力、圧力、振動などの生体外部からの物理的・化学的・刺激についても、これらの刺激を仮想的な生体分子として扱い、対応する標的・生体分子との生体分子対を定義してもよい。

「構造コード」とは、生体分子がDNAか、RNAか、蛋白質か、ペプチドか、一般低分子か等の構造上の特徴を表す分類コードである。

「機能コード」とは、生体分子の分子レベルでの機能を表わす分類コードであり、例えば、「構造コード」が「蛋白質」である生体分子の場合は、膜受容体・核内受容体・トランスポーター・メディエーター・加水分解酵素・リン酸化酵素・

脱リン酸化酵素等の分類を表わし、「構造コード」が低分子である生体分子の場合は、基質・生成物・前駆体・活性ペプチド・代謝産物等の分類を表わす。

「関係コード」とは、生体分子対をなす2分子間の関係を表わす分類コードである。例えば、アゴニストと受容体であれば10，酵素と基質であれば21，基質と生成物であれば22というように類型化することができる。例えばtwo-hybrid 実験法で相互作用があるとされた2ヶの蛋白分子のように2分子の相互の役割が明確でない場合は、その旨を区別したコードを用いることが望ましい。

「関係機能コード」とは、生体分子対をなす2分子の直接結合に伴う現象や変化を表わす分類コードであり、例えば加水分解、リン酸化、脱リン酸化、活性化、不活化等の分類を用いる。

「信頼性コード」とは、生体分子対毎に直接結合の信頼性のレベルや直接結合の根拠となった実験法などを示すためのコードである。

「コネクト検索」とは、1以上の任意の生体分子又は生体イベントを指定して、それらを含む機能上又は生合成上関連した分子のつながりを自動的に探すことをいう。

「分子機能ネットワーク」とは、生体分子連鎖データベースを用い、1以上の任意の生体分子又は生体イベントを指定してコネクト検索した結果得られる、機能上又は生合成上関連した分子のつながりをいう。

「医薬分子」とは、医薬として製造され治療に用いられる化合物の分子を指すが、医学・薬学研究用に用いられる化合物や特許明細書や文献に記載の化合物など生理活性が既知の化合物も含むものとする。

「生体イベントの情報と関係付ける」とは、ある生体イベントの発現に、ある生体分子又は医薬分子又は遺伝子情報又は分子機能ネットワークが関わっていることを示すか又は見つけることをいう。

「パターン化」とは、生体分子、生体分子対、生体イベント等に関する情報をデータベースに記録する際に、与えられた情報をそのまま記録するのではなく、

予め決められたカテゴリーに該情報を分類し、該当するカテゴリーを表わす記号により該情報を表現することをいう。上記の「構造コード」、「機能コード」、「関係コード」、「関係機能コード」について挙げられた例が、「パターン化」の例である。

「生成臓器」とは、生体分子が生成される臓器、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などをいう。

「存在臓器」とは、生体分子が生成された後に蓄えられる臓器、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などをいう。

「作用臓器」とは、生体分子又はキイ分子が生体イベントを発現する臓器、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などをいう。

本発明の実施の一つの態様として以下のような方法が提供される(第1図)。まず、直接結合する2ヶの生体分子の対に関する情報を蓄積した「生体分子連鎖データベース」を作成する。生体分子の分子略号の付け方など、生体分子自身に関する情報をここに含めてもよいが、別データベースである「生体分子情報データベース」で行うのが望ましい。次に、上記の「生体分子連鎖データベース」から、1以上の任意の分子を指定してコネクト検索を行ない、1以上の生体分子の機能上又は生合成上のつながりの表現である「分子機能ネットワーク」を得る。

生体分子対のうち、少なくともキイ分子とその標的生体分子からなる生体分子対に対して生体イベントの情報を関連付けておくことにより、「分子機能ネットワーク」とともに、分子機能ネットワーク中の分子が直接又は間接に関わる生体イベントを推定することができる。さらに、キイ分子の量的又は質的な変動と生体イベントのアップオアダウンの関係の情報を付加しておくことにより、分子機能ネットワーク上の任意の分子の量的又は質的な変動が生体イベントの亢進・上昇に働くか、抑制・低下に働くか等を推定できる。

「生体分子情報データベース」の主たる役割は、各生体分子の正式名称に対して分子略号又はIDを定義することであり、その他にも生体分子自身についての

必要な情報をまとめて保存するとよい。例えば、分子名称、分子略号、構造コード、機能コード、生物種、生成臓器、存在臓器等に関する情報を保存するのが望ましい。また、実験的に単離や存在の確認がされていない生体分子であっても、例えば他の生物種での実験から存在が推定される分子に仮に分子略号その他の情報を与えて定義してもよい。

「生体分子情報データベース」には、各生体分子のアミノ酸配列や構造に関する情報を含めてもよいが、該情報を配列データベースや構造データベースに別途保存し、分子略号に基づいて必要に応じて取り出すのが望ましい。生体分子のうち低分子量のものについては、必要に応じて分子機能ネットワークの表示に化学構造を加えることができるように、正式な分子名称だけでなく、化学構造の表記に必要なデータを生体分子情報データベース又は別途のデータベースに保存しておくことが望ましい。

2以上の生体分子が多量体又は集合体として活性を示す、或いは機能するなど、複数の生体分子をまとめて扱う方が便利な場合には、それらを1ヶの仮想的生体分子として定義し、分子略号を付して「生体分子情報データベース」に登録してもよい。その場合、その構成分子が既知の場合には、それぞれに分子略号を付して登録しておき、仮想的生体分子のレコードに構成分子の分子略号を記述するフィールドを設けるとよい。どのような生体分子で構成されているか不明な場合でも、集団として特定の機能をもつ仮想的生体分子を定義して、生体分子対の定義に用いることが可能である。

また、生体分子が2以上のドメイン構造からなり、かつそれらが互いに異なる機能を持つなどの理由で各ドメインを独立に扱う方が有利と判断された場合には、各ドメインを独立の分子として扱ってもよい。例えば、元の生体分子と共に、各ドメインに分子略号を付して生体分子情報データベースに登録するとよい。元の生体分子のレコードに、分割したドメインの分子略号を記述するフィールドを設けることで、1ヶの生体分子が2以上の異なる機能をもつことが記述できる。遺伝子ではないゲノム配列上の特定の配列が、ある機能を持つ場合や特定の生体分

子によって認識される場合には、その配列の部分で独立の生体分子として扱い、分子略号を付して生体分子対の定義に用いることができる。

生体分子対についての情報は「生体分子連鎖データベース」に保存する。生体分子対毎に、対をなす2つの生体分子の分子略号、関係コード、関係機能コード、信頼性コード、生体イベント、作用臓器、共役分子、その他の付加情報等を収録しておく。キイ分子とその標的分子の分子対については、生体イベント、両分子の量的又は質的な変動に対応した生体イベントのアップオアダウンの情報、病態イベント等の情報をできる限り入力しておくことが望ましい。キイ分子以外の生体分子対についても、該生体分子対が発現に直接関わる生体イベント又は病態イベントがある場合には、その生体イベントや病態イベントを入力しておくのが望ましい。キイ分子の量的又は質的な変動に対応した生体イベントのアップオアダウンの情報としては、例えば、キイ分子の増加に対して生体イベントが正常な範囲に比べて亢進するか低下するか等の単純化した情報で示してもよい。1ヶの酵素が2種以上の基質の反応を触媒してそれぞれ異なる反応生成物を生成する場合には、酵素と基質と反応生成物の関係を特定するための表現を付け加える。

「生体分子情報データベース」と「生体分子連鎖データベース」は、その内容や構成が異なるため、本明細書では概念上独立のデータベースとして扱っているが、本発明の趣旨からも、双方を合わせた1ヶのデータベースに2種類のデータを含めても良いのは言うまでもない。また、「生体分子情報データベース」と「生体分子連鎖データベース」はそれぞれ2以上存在していてもよく、この場合、各データベースを適宜選択又は組み合わせて用いることができる。例えば、特定のフィールドで区別した、異なる生物種についてのデータを同じ「生体分子情報データベース」と「生体分子連鎖データベース」に保存してもいいし、ヒトとマウスについて別々のデータベースを作成して保存してもかまわない。

「関係コード」としては、該生体分子対を構成する2分子がアゴニストと受容体である、酵素と基質である、というように単語で入力してもよいが、アゴニストと受容体の関係なら10、酵素と基質の関係なら21、酵素と生成物の関係な

ら 2 2 , といった具合に類型化して入力するのが望ましい。また、「関係機能コード」として、加水分解、リン酸化、脱リン酸化、活性化、不活化等の機能の別を保存しておくとし便利であるが、これも類型化して入力しておくのが望ましい。

酵素と基質のように互いの関係が明確になっている場合だけでなく、例えば two-hybrid 実験法で蛋白-蛋白相互作用があるとされた 2 ケの蛋白分子のように、両者の相互の役割が明確でない場合がある。このような生体分子対を含めてコネクト検索を行うためには、生体分子対をなす 2 分子の関係に方向性があるかないかを区別して扱うと便利である。各生体分子対に対して、どちらのケースに属するかを区別できるような関係コードを用いるのが望ましい。前者の場合は、作用方向確定として分子対の表現における 2 分子の入力順序のみが検索時に考慮されるのに対して、後者の場合は、検索時に作用方向不明として逆方向の関係も考慮される。

直接結合する生体分子対の情報も、実験的に確実に証明されているものから、とりあえず生体分子対であると仮定されているものまで様々ある。また、実験法によっては false positive のために間違っ生体分子対であるとされる場合も含まれる。そこで、各生体分子対の情報に信頼性のレベルや実験法を示す「信頼性コード」を付加しておくのが望ましい。検索により生成する分子機能ネットワークが多すぎる場合に、このコードを用いて、絞り込むことができる。

生体分子が生成される臓器とは別に、それが蓄積され存在する臓器及び作用する臓器の情報を保持しておくとし、分子機能ネットワークの生成時に、例えばある臓器で生成されて細胞外に出た分子が他の細胞の膜上にある標的生体分子に細胞外から作用する、といった現象を容易に表現できる。生体分子の生成臓器と存在臓器の情報は「生体分子情報データベース」に、作用臓器の情報は「生体分子連鎖データベース」に入力しておくのが望ましい。ここで生成臓器・存在臓器・作用臓器の記述は臓器に限定される必要はなく、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などの情報を含んでいてもよい。

直接結合することを実証した実験や推定の方法、生体イベントの種類、キイ分

子の量的変動に対応した生体イベントのアップオアダウン、細胞内部位や組織、臓器、臓器内部位の表現は単純化してあれば何でもよいが、類型化して短い英数字記号等に変換しておくのが望ましい。同義語辞書で定義しておけば、同時に同義語の処理もでき、入力時のミスが最小化できるからである。

以下に「生体分子連鎖データベース」から「分子機能ネットワーク」を生成する「コネクト検索」の概念を示す。本発明の「コネクト検索」には、この概念を実現できるものであれば、いかなる方法を用いてもよい。例えば、Sedgewick による「アルゴリズム C (近代科学社、1996)」第 29 章に記述されている「深さ優先探索」のアルゴリズムなどが利用できる。

分子略号 a~z で表される生体分子からなる各生体分子対を(n,m) のように表現するとすると、生体分子連鎖データベースは次のような生体分子対の集合として表される。

(a, c) (a, g) (b, f) (b, k) (c, j) (c, r) (d, v) (d, y) (e, k) (e, s)
(g, u) (j, p) (k, t) (k, y) (p, q) (p, y) (x, z)

コネクト検索で、例えば c と e を含む分子機能ネットワークを生成するよう指定したとすると、対のうち的一方の分子を共通にもつ生体分子対 (c, j) (j, p) (p, y) (y, k) (k, e) を次々に探し、

分子機能ネットワークとして、分子 c, j, p, y, k, e のつながりである

c - j - p - y - k - e を得る。

得られた「分子機能ネットワーク」に基づいて、生体イベントの推定を以下のようにして行う事ができる。生体分子 e がキイ分子であって生体イベント E の情報をもつ場合、生体分子 c, j, p, y, k は生体イベント E の発現に直接又は間接に関係すると推定できる。さらに、例えば分子 e が減少すると E の発現が亢進するといった生体イベントのアップオアダウンの情報がある場合には、(c, j) (j, p) (p, y) (y, k) (k, e) のそれぞれの関係を考慮して、c, j, p, y, k のうちの任意の分子の量的又は質的変動に対する生体イベント E の発現への影響を推定することができる。

さらに、ある生体分子からキイ分子までの分子機能ネットワーク上にある N 個の生体分子が、生体イベントの発現量 Q_E に与える影響を、例えば次のような式で予測することもできる。ここで、 S_i は i 番目の生体分子の状態の質的な評価値、 R_i は i 番目の生体分子の量を表わす値、 V_i は i 番目の生体分子が存在する環境の評価値を表わし、 f は $3 \times N$ 個の入力値をもつ多価関数である。

$$Q_E = f (S_1, R_1, V_1, \dots, S_N, R_N, V_N)$$

1 個の分子機能ネットワークが関係する生体イベントは 1 種類に限らないし、また、1 種類の生体イベントに関わる分子機能ネットワークが数個あることが予想されるが、生体イベントの側から関連のある分子機能ネットワークを絞り込むことは可能である。例えば、1 以上の生体分子を指定して「分子機能ネットワーク」を生成した場合に、著しく多数の生体分子が含まれる「分子機能ネットワーク」が生成された場合には、生体イベントの情報を加えて「分子機能ネットワーク」の範囲を絞り込むことが可能である。当然のことながら、何らかのメディエーター分子、又は該分子と標的生体分子の関係を含むことを条件として、「分子機能ネットワーク」を生成させることも可能である。

また、「生体分子連鎖データベース」のデータを適宜分割し、又はフィルターをかけ、又は部分集合を抽出し、又は階層化することによって、必要な範囲の分子機能ネットワークを生成することができる。分割やフィルターや部分集合の抽出は、本発明のデータベース特有のデータ項目に対する検索、キーワードを用いた一般的な文字列検索、アミノ酸配列又は核酸配列に対する相同性検索、化学構造式に対する部分構造検索等の検索方法により行なうことができる。これらの検索を予め「生体分子連鎖データベース」又は「生体分子情報データベース」に対して行なうことにより、限定をかけた分子機能ネットワークや特徴づけした分子機能ネットワークを生成することができる。例えば、生成臓器や作用臓器の情報を用いて、肝臓で生成する生体分子とか皮膚で起こる生体イベントといった観点から絞り込んだ部分データベースを作成してコネクト検索することで、限定した範囲の「分子機能ネットワーク」を生成することができる。また、コネクト検索で

生成させた分子機能ネットワークに対して、そこに含まれる生体分子又は生体分子対について上記の検索を行なって分割やフィルターや部分集合の抽出をすることによって、所望の特徴をもつ分子機能ネットワークや所望の範囲の分子機能ネットワークを生成することができる。こうした限定や特徴づけは検索を容易にするだけでなく、特定の生体分子群や生体分子対を分子機能ネットワーク上で強調して表示して、分子機能ネットワークの理解を助ける目的にも有効である。

「生体分子連鎖データベース」の分割やフィルターや部分集合の抽出を、ネットワーク上の連なりに基づいて適切に行ない、その包含関係を示す情報を保存して利用することにより、「分子機能ネットワーク」の階層化が可能になる。未解明の分子や分子間のつながりが一部にあっても、それらをまとめて1ヶの仮想的な生体分子として他分子との対を定義し、仮の分子機能ネットワークを生成することも可能である。含まれる分子数が多いために過度に複雑なネットワークが生成される場合には、ネットワーク上で連結している2以上の生体分子群を仮想的な1ヶの生体分子と定義して、ネットワークをシンプルに表現することも可能である。

このような階層化を利用することにより、コネクト検索の高速化を可能にし、ネットワーク表示の詳細さを調節可能にして、過度な複雑さを適切に回避できる。本明細書中では、このようにネットワーク上で連結している2以上の生体分子対からなる部分的なネットワークを「サブネット」とよぶ。

サブネットとしてはいかなる部分ネットワークを指定してもよいが、好ましくは代謝系におけるTCA回路やペントースリン酸回路の例のように研究者によく知られたカスケード、パスウェイ、回路等をサブネットとするのが便利である。また、あるサブネットが別のサブネットに含まれていてもよく、例えば代謝系自体を複数のサブネットを含む上位のサブネットと見なすこともできる。

各サブネットを1ヶの仮想生体分子として扱う方法もあるが、サブネットを構成する生体分子対とサブネットの階層に関する情報を「生体分子連鎖データベース」に保存するのが便利である。また、「生体分子連鎖データベース」にサブネッ

トを表わすための上位のデータ階層を設けて、そこに該サブネットについての情報を保存してもよい。生体分子対のサブネットによる階層化は2層に限らず、複数のサブネットの集合をより上位のサブネットとして保存してもよい。分子機能ネットワーク生成時に個々の分子対のデータと上位階層のサブネットデータの間での相互参照を容易にするために、個々の分子対データ・サブネットデータそれぞれに互いの関連を示す情報を格納しておくのが望ましい。1ヶの生体分子対が複数のサブネットに関連付けられてもよいことは言うまでもない。

階層化した「生体分子連鎖データベース」のサブネットのデータには、下位階層の生体分子対との関連のみならず、サブネット間の関連の情報を含めるのが望ましい。例えば、代謝系において解糖系とTCA回路は連続して動作するサブネットであり、これらのサブネット間の関係を上位階層における対の関係として保存することができる。この場合、サブネットの対の情報に加えて、サブネット間の接点となる生体分子の情報を加えておくことが望ましい。

さらに、ネットワークの階層化に加えて、生体分子自身も階層化でき、「生体分子情報データベース」にその情報を保存して利用できるのが本発明の特徴である。ネットワークの高速な検索と便利で多様な表示のためには、生体分子情報と生体分子対情報の双方を階層化しておくのが望ましい。生体分子の階層化の対象としては、以下のような例を挙げることができる。生体分子のなかには、複数の異なる分子が特異的に集合してある機能を示すものがあり、分子の集合状態の違いにより機能の発現状態や種類が制御されている場合も多い。また、免疫細胞などで見られるように、細胞表面に発現している複数の分子の組み合わせにより、生体イベントとの関連や細胞の機能が規定されている場合もある。このような場合に、上記のように分子の集合状態を1ヶの仮想的生体分子として扱う方法もあるが、別の方法として、「生体分子情報データベース」に分子の集合状態を表わすための上位のデータ階層を設けて、そこに該集合状態についての情報を格納してもよい。分子機能ネットワーク生成時に生体分子データと上位階層のデータの間での相互参照を容易にするために、生体分子データと上位階層データのそれぞれに、互い

の関連を表わす情報を格納しておくのが望ましい。1ヶの生体分子が複数の上位階層データに関連付けられてもよいことは言うまでもない。

生体イベント・病態イベントのなかには、特定の生体分子対に関連付けすることのできないものも多い。例えば、生体イベント・病態イベントとあるサブネットの形成との間の関係は分かっているが、該イベントに直結する生体分子対が未知の場合がある。このような場合には、上記の生体分子対データの階層化を利用して、生体イベント・病態イベントを生体分子対の上位階層であるサブネットのデータに対して関連付けることで、該イベントと生体分子ネットワークの関係を記述することが可能となる。

また、特定の分子の集合状態や細胞表面への特定の分子の発現状態がある生体イベント・病態イベントの発現に関連している場合には、上記の分子集合状態や分子発現状態の階層化を利用して、生体イベント・病態イベントを分子集合状態や分子発現状態の階層のデータに関連付けておくことで、該イベントと生体分子ネットワークの関係を記述することが可能となる。

さらに、生体イベント・病態イベントのなかには、特定の生体分子対やサブネットのいずれにも関連付けすることができないものもある。このような例の一つとして、炎症性サイトカインの遊離・白血球の組織への浸潤・毛細血管の透過性の向上などの様々な生体イベントが組み合わさって起こる「炎症」という病態イベントを挙げることができる。このようなイベントを扱うためには、生体イベント・病態イベントを階層化し、下位の階層には生体分子対やサブネットに関連付けられるイベントを記述し、上位階層には下位階層のイベントに関連して起こるイベントを記述するとよい。この階層化では、2層以上の階層構造を用いてもよいことは言うまでもない。各階層間でのイベントの相互参照を容易にするために、各階層のイベントのデータに上下の階層のデータへの関連を示す情報を格納しておくのが望ましい。このような生体イベント・病態イベントのデータの階層化により、特定の生体分子対やサブネットに直接関連付けられないイベントについても、分子機能ネットワークとの関係を記述することが可能となる。

以上例示したように「生体分子情報データベース」及び「生体分子連鎖データベース」のデータを階層化して保存することにより、多様な用途に対応した分子機能ネットワークの生成を効率的に行なえるようになる。

解糖系に存在するある生体分子（分子A）とあるキナーゼカスケードに存在するある蛋白（分子B）との関連を調べようとする場合に、階層化していないデータを用いる方法では膨大な数の分子対を対象にコネクト検索を行なう必要があり、分子Aと分子Bの間の経路が長い場合には事実上検索が不可能となる。一方、階層化されたデータを用いると、「解糖系」というサブネットと「あるキナーゼカスケード」というサブネットとの間の関連をサブネットという上位階層においてコネクト検索し、上位階層での経路が見つかった場合には、必要に応じて経路上にある各サブネットの下位階層でコネクト検索を行なうことができる。このように経路探索問題を異なる階層の問題に分割することにより、階層化を用いない場合には不可能であったような分子機能ネットワークの生成も可能となる。

また、上記の階層化データを利用したコネクト検索において、特定のサブネットが頻繁に参照されるような場合には、該サブネット内について予めコネクト検索を行なっておき、該サブネット内の分子機能ネットワークの情報を保存しておくといふ。この処理により、全体の分子機能ネットワークの生成をより効率的に行なえるようになる。

さらにまた、例えば、「炎症」という病態イベントに関わる分子機能ネットワークを生成する場合に、「炎症」という上位階層のイベントに関連する下位階層のイベントを探し、該下位階層イベントが関連する生体分子対又はサブネットから出発してコネクト検索することにより、より広範囲な分子機能ネットワークの生成が可能となる。

上記のように、本発明によると生体分子間の直接結合の情報の情報から、任意の分子に関連した分子機能ネットワークが生成でき、直接又は間接に関わる生体イベント及び病態イベントが容易に推定できる。また、本発明は逆に、疾患で特徴的に現れる生体イベントや病態イベント、生体分子の量の変化などの知見から、

疾患に関連する可能性の高い分子機能ネットワークを選別し、疾患の分子的メカニズムを推定する目的に利用できる。さらに、本発明によると、特定の疾患や症状の治療にネットワークのどの過程を阻害するのが有効か、ネットワーク中のどの分子が創薬ターゲット（医薬開発で標的とする蛋白質その他の生体分子）として有望か、その創薬ターゲットから予想される副作用は何か、それを回避するにはどのようなアッセイ系で医薬開発候補を選別するかと言った創薬戦略の構築が可能になる。

医薬分子は一般に、体内で蛋白質をはじめとする生体高分子に結合してその機能を制御することにより薬理活性を発現する。それらの分子の作用は生体分子の作用よりも詳しく研究され、対象疾患の分子メカニズムの解明に役立ってきた。そこで、製造承認を受けて医療に用いられている医薬分子や薬理学研究等に用いられる薬物分子とその標的分子の対の関係を、上記の生体分子及び生体分子間の情報に加えることにより、本発明の方法の有用性を高めることができることに注目した。標的分子は殆どの場合、蛋白質又は糖などで修飾された蛋白質である。標的分子を含む分子機能ネットワークから、副作用となりそうな生体イベントを推定することができるし、併用する薬剤の関わる分子機能ネットワークとの交差から、薬剤間相互作用を推定することも可能になる。その結果、副作用リスクや薬剤間相互作用のリスクを考慮した薬剤の選択や用量の設定が可能になる。

以下に本発明で医薬分子と標的分子の対の関係を加えた場合の方法の例を示す。医薬分子の正式名称に対して、分子略号を定義し、該分子そのものについてのすべての情報を収録する「医薬分子情報データベース」を作成する。ここには医薬分子の名称、分子略号、適用疾患、用量、標的分子その他の情報を保存する。生体分子情報データベースの場合と同様に、医薬分子の化学構造、アミノ酸配列（ペプチド又は蛋白質の場合）、立体構造等の情報を「医薬分子情報データベース」に含めてもよいが、別途のデータベースに保存することが望ましい。医薬分子と生体分子、蛋白質と低分子などを区別する目的には、構造コード等で

区別してもよいし、最初の1文字で区別できるような分子略号の付け方を採用してもよい。さらに、医薬品の添付文書やその他の文献から、顕著な副作用、他の薬剤との相互作用、代謝酵素等の情報も入力しておくこと、分子機能ネットワークに基づいて、遺伝子多型との関連で薬剤の適正な選択等を行う目的に役立つ。

さらに、医薬分子と標的蛋白質の対とその関係に関する情報を含むデータベースである「医薬分子連鎖データベース」を作成する。医薬分子の分子略号、標的生体分子の分子略号、関係コード、薬理作用、適用疾患その他の情報を保存する。標的生体分子の分子略号については、生体分子情報データベース中で定義したものをを用いる必要がある。関係コードなど生体分子連鎖データベースと共通するデータ項目については、生体分子連鎖データベースに準じた表記方法を用いるとよい。

「医薬分子情報データベース」と「医薬分子連鎖データベース」を作成して医薬分子及び医薬分子対の情報を取り込むことにより、第2図に示すように本発明の方法を拡張することができる。ここでのコネクト検索による分子機能ネットワークの生成や生体イベントの推定などは、上記の生体分子連鎖データベースと生体分子情報データベースだけを用いた場合と同様の方法で行うことができ、該ネットワーク上の分子を標的とする既存医薬分子の情報が同時に得られる。また、生体分子連鎖データベースと生体分子情報データベースだけを用いて作成した分子機能ネットワークから、指定した医薬分子の関わる分子機能ネットワークを抽出する目的にも利用できる。

一方、ヒトゲノム配列の解析をはじめとして、多様な角度から遺伝情報の解明が急速に進みつつある。ゲノムワイドにcDNAが単離され、o r f (open reading frame) や遺伝子配列の解明が進み、各遺伝子のゲノム上への位置づけが進んでいる。そこで、本発明の別の態様として、生体分子のうちの蛋白質の分子略号と、該蛋白質をコードする遺伝子の名称、略称、IDその他の情報を関係付けた生体分子-遺伝子データベースを作成しておくことにより、以下のように本発明の方法を拡張することができる。すなわち、遺伝子と生体分子を対応づけること

により、疾患のマーカーとなる遺伝子や蛋白の意味付け、疾患と遺伝子多型の関わり等の知見を分子機能ネットワーク上の分子及び生体イベントとの関連において理解することが可能になる。生体分子-遺伝子データベースには、生物種、ゲノム上の位置、遺伝子配列、機能のほか、遺伝子多型のアミノ酸変異と略称、機能との関連等の情報を含めておくのが望ましく、必要に応じて2以上のデータベースとするのもよい。

ゲノム配列上に位置づけられた遺伝子名や遺伝子の並びから、特定のキイ分子が核内受容体に作用することによって転写される蛋白が明確になり、生体分子間の相互の制御の関係を分子機能ネットワークに反映することが可能になる。また、臓器によって発現する遺伝子や蛋白が異なることが知られているが、本発明の方法によれば、そうした発現情報を「生体分子情報データベース」に取り込むことにより、臓器毎に異なる「分子機能ネットワーク」を生成することができ、例えば、異なる臓器では核内受容体を標的とする医薬分子の作用が違ったり逆転したりする現象の説明が可能になる。また、医薬分子を投与した場合に蛋白発現が変化する事実がわかっているが、本発明の方法により個々の発現蛋白量の増減を標的的生体分子に関わる分子機能ネットワーク上で解釈すると、遺伝子多型を考慮した薬剤選択に有用である。

上記の遺伝子転写や蛋白発現の情報の保存においても、階層化の概念を利用することにより、より効率的かつ広範囲な分子機能ネットワークの生成が行なえるようになる。例えば、特定の核内受容体により転写・発現される複数の遺伝子・蛋白に対して、「生体分子情報データベース」に遺伝子群の転写・蛋白群の発現を表す上位の階層を設けて、そこに該遺伝子群・該蛋白群のデータを保存しておくといふ。該遺伝子群の転写や該蛋白群の発現に関連する生体イベント・病態イベントが存在する場合には、「生体分子連鎖データベース」に該遺伝子群・該蛋白群の上位階層データと該イベントとの関連を記載することにより、個々の遺伝子・分子と該イベントとの間の関連付けでは表せないような分子機能ネットワークの生成が可能となる。

上記の遺伝子転写や蛋白発現の情報の階層化した保存法において、該遺伝子群の個々の遺伝子又は該蛋白群の個々の蛋白の転写又は発現の量的な情報が分かっている場合には、それらの情報も数値パラメタとして「生体分子情報データベース」に保存しておくといよい。これらの数値パラメタを利用することにより、個々の遺伝子の転写量又は個々の蛋白の発現量の違いに応じて、関連する生体イベント・病態イベントが変化する場合を記述することが可能となる。

また、ゲノムや遺伝子についての個体の多様性が明らかにされつつあるが、そうした情報を本発明の方法とリンクすることにより、個体差への理解が進み、個体差に基づく治療が可能になる。特定の生体分子（蛋白質）の機能が損なわれるような遺伝子多型について、分子機能ネットワーク上で解釈することにより生体イベントへの影響が推定できる。1ヶの遺伝子の欠損や異常が引き起こす遺伝病の症状及び生体イベントの異常の情報を、本発明の方法にリンクさせることも、その理解に有用である。

いくつかの代表的疾患で、各疾患の患者に頻度高く見出される数ヶの遺伝子、疾患背景遺伝子の存在が報告されている。特定の疾患にかかりやすい遺伝的体質が実際にあると仮定した場合、例えば、血圧の調節に関わる分子機能ネットワークは2以上あり、そのうちのどのネットワークのどの分子の異常かによって、高血圧の背景となる遺伝子は相当数あっても不思議ではない。こうした polygenic な遺伝子の問題を解釈するには、本発明の方法が不可欠である。

また、近年、マウス、ラット等の動物のゲノムや遺伝子の解析も急速に進められており、ヒトのゲノムや遺伝子との対応も付けられるようになってきた。これらの動物とヒトとで生理機能の調節に関わる蛋白がかなり似ていることが期待されるが、かなりの違いもあることが医薬開発の障害となってきた。これらの動物とヒトの間で蛋白や蛋白機能が大きく異なる場合があることがわかってきており、本発明の方法とのリンクによって、ヒトの分子機能ネットワークとの違いを明らかにすることは医薬開発に有用である。さらに、一般にヒト用に開発された医薬を転用することが多い動物薬について、適切な使用を目指す上でも有用である。

医薬開発においてはヒトの疾患と病態が類似した病態動物がある場合、その動物での薬理活性を指標に開発を進めることが多い。そうした病態動物の遺伝子も進められており、本発明の方法によりヒトの遺伝情報と対応づけることによって、ヒトの該疾患のメカニズムの解明に役立つ。

さらに、遺伝子機能の解明の目的で、特定の遺伝子を潰したノックアウト動物や機能の弱い遺伝子に変えたり遺伝子を過剰発現させたりしたトランスジェニック動物を作成することが多くなった。これらは致死的で生まれて来ない場合や生理機能や行動に何の影響も見られない場合も多いが、生まれてきた動物に何らかの異常が観察される場合でも、これらの動物実験の結果の解析は非常に難しいといわれている。このような実験において、本発明の方法を用いて、該遺伝子操作の影響を予測してから、機能解析すると便利である。

遺伝子関連の情報は、解析の進展とともに、配列 I D の面からの統合の試みも進みつつあり、さらにゲノム配列上に遺伝子を位置づける試みも進められている。上記「生体分子連鎖データベース」との連携を考慮した独自の遺伝情報データベースを構築して上記の目的に利用することも考えられるが、これらの情報が膨大でかつ公開される方向にあることを考えると、将来はそうした公開の情報を本発明の方法に臨機に取り込むことによって、上記の方法が実施できる可能性が高い（第 3 図）。

本発明の方法で用いる生体分子連鎖データベースは、必ずしも同一の場所にまとめて管理・保存されている必要はなく、分子略号を統一することにより、異なる場所に管理・保存されている 1 以上の生体分子連鎖データベースを適宜選択して通信等の手段により接続して用いることもできる。生体分子連鎖データベースのみならず、本発明の方法で用いる生体分子情報データベース、医薬分子連鎖データベース、医薬分子情報データベース、遺伝情報データベースなどについても、同様の扱いが可能なことは言うまでもない。

本発明の実施のさらに別の態様として、生体イベントの発現に直接関わる生体分子と該生体イベントの情報を含むデータベース（生体イベントー生体分子デー

データベース)を作成し、生体イベントの情報を必ずしも含まない分子ネットワークのデータベースと併用する方法も提供される。さらに別の実施の形態として、生体イベントの情報を必ずしも含まない分子ネットワークのデータベースから、任意の分子に関わる部分的な分子ネットワークを抽出し、該ネットワークを構成する分子に基づいて、上記の生体イベント-生体分子データベースを検索する方法も提供される。

本発明の実施のさらに別の形態として、「生体分子情報データベース」・「生体連鎖データベース」・「医薬分子情報データベース」・「医薬分子連鎖データベース」・「生体分子-遺伝子データベース」などの各データ項目に対して、キーワード・数値パラメタ・分子構造・アミノ酸配列・塩基配列などに基づく検索を行ない、該検索結果に基づいて分子機能ネットワークを生成する方法が提供される。以下に、検索に基づく分子機能ネットワーク生成の例を挙げるが、本発明の範囲がこれらの例に限定されるものでないことは言うまでもない。

各データベースには、分子名称・分子略号・生物種・生成臓器・存在臓器等の種々の情報がテキストとして保存されている。これらのテキストに対して文字列の全一致・部分一致などに基づく検索を行なうことで、生体分子・生体分子対・生体イベント・病態イベント・医薬分子・医薬分子-生体分子対・遺伝子-蛋白の対応データなどを絞り込むことができる。これらの絞り込まれた情報に基づいて、コネクタ検索の起点・終点を定めたり、コネクタ検索で対象とする分子対の範囲を狭めたりすることができ、利用目的に応じた分子機能ネットワークの生成が可能となる。

「医薬分子情報データベース」に医薬分子の化学構造・立体構造が保存されている場合には、これらに対して全体構造の一致・部分構造の一致・構造類似性などに基づく検索を行なうことで、医薬分子を絞り込むことができる。絞り込まれた医薬分子に基づいて、該医薬分子に関連する分子機能ネットワークを生成したり、該医薬分子に関連する生体イベント・病態イベントを検索したりすることが可能となる。

「生体分子情報データベース」に遺伝子転写・蛋白発現等の数値パラメタが保存されている場合には、これらの数値パラメタに基づく検索を行なうことで、遺伝子の転写量・蛋白発現量等に応じた分子機能ネットワークの生成が可能となる。

「生体分子情報データベース」又は関連データベースに蛋白のアミノ酸配列が保存されている場合には、これらのアミノ酸配列に対して配列相同性や部分配列パターンの一致などに基づく検索を行なうことで、生体分子を絞り込むことができ、該生体分子に基づいて分子機能ネットワークを生成することができる。この方法は、機能未知の蛋白質やその部分配列情報について、該蛋白質が関係する可能性が高い分子機能ネットワークを推定し、さらに該蛋白質の機能を推定する目的に有効である。

「生体分子情報データベース」、「生体分子－遺伝子データベース」又は関連データベースに蛋白に対応する遺伝子の塩基配列が保存されている場合には、これらの塩基配列に対して配列相同性や部分配列パターンの一致などに基づく検索を行なうことで、生体分子を絞り込むことができ、該生体分子に基づいて分子機能ネットワークを生成することができる。この方法は、機能未知の遺伝子やその部分配列情報について、該遺伝子から翻訳される蛋白が関係する可能性が高い分子機能ネットワークを推定し、さらに該蛋白の機能を推定する目的に有効である。

本発明の実施のさらに別の態様として、本発明の方法を実施するためのプログラムとデータベースからなるコンピュータシステム；本発明の方法を実施するためのプログラムとデータベースを記録したコンピュータ読み取り可能な媒体；本発明の方法で用いるためのデータベースを記録したコンピュータ読み取り可能な媒体；本発明の方法により生成された分子機能ネットワークに関する情報を記録したコンピュータ読み取り可能な媒体；なども提供される。

本発明の方法の特徴をまとめると、以下の通りである。

- ・生体イベントの情報を含み、直接結合する生体分子対の情報を蓄積することにより、生体内の分子間の関わりデータベースを作成する。
- ・パーツの集合である上記データベースからコネクト検索で、任意の生体分子又

は生体イベントに関わる分子機能ネットワークを作成する。

- ・分子機能ネットワークに基づいて、任意の分子が直接間接に関わる生体イベントを推定する。
- ・生体イベントの情報つき分子機能ネットワークから、疾患メカニズム、可能な創薬ターゲット、副作用リスク等を推定する。
- ・生体分子の量的又は質的な変動から生体イベントのアップオアダウンを推定する。
- ・生体分子の生成臓器、存在臓器及び作用臓器の情報をもつ分子機能ネットワーク
- ・医薬分子情報と分子機能ネットワークを用いた副作用、薬剤間相互作用の推定
- ・医薬分子の投与による蛋白発現の変化を分子機能ネットワーク上で解釈
- ・遺伝情報とのリンクによって分子機能ネットワークへの遺伝子多型の影響、疾患背景遺伝子等の解析を行う。

実施例

以下、本発明を実施例によりさらに具体的に説明するが、本発明の範囲はこれにより限定されるものではない。

実施例 1

レニン-アンジオテンシン系について分子機能ネットワークの生成を試みた例を示す。レニン-アンジオテンシン系は、生体で血圧を調整する主要な機構の一つであり、関与する生体分子の多くが明らかにされている(第4図)。これまでに知られているレニン-アンジオテンシン系に関与する生体分子について、生体分子情報データベース(第5図)と生体分子連鎖データベース(第6図)を作成し、生体分子と生体イベントを質問として与えることにより、分子機能ネットワークの生成を試みた。

第7図に、生体分子の一つである「アンジオテンシンⅠ」と生体イベントの一

種である「血圧上昇」を質問として与えて生成された分子機能ネットワークを示す。生体分子連鎖データベースをコネクト検索することにより、「アンジオテンシンⅠ」から「血圧上昇」に至るまでに関係する生体分子と、それらにより作られる分子機能ネットワークが得られる。

さらに、血圧降下作用をもつ医薬分子に関する医薬分子情報データベース（第8図）と医薬分子連鎖データベース（第9図）を作成し、生体分子情報データベース（第5図）と生体分子連鎖データベース（第6図）と併せて用いることにより、医薬分子が関わる分子機能ネットワークの生成を試みた。

第10図に、医薬分子の一つであるエナラプリルと生体イベントの一種である「血圧上昇」を質問として与えて生成された分子機能ネットワークを示す。エナラプリルはそれが直接結合するアンジオテンシン変換酵素に対して阻害剤の関係にあるため、アンジオテンシン変換酵素が直接結合する関係（酵素－基質の関係）であるアンジオテンシンⅡへの連鎖が断たれ、それ以降のネットワーク上に存在する「血圧上昇」のイベントが抑止（停止）されることが示される。

実施例2

本発明を分子機能ネットワークの検索及び表示のためのプログラムとして実現した例を示す。第11図に本実施例の検索及び表示の流れ図を示すが、これらの工程は本発明をプログラムとして実現する一つの例を示したものに過ぎず、本発明の範囲はこの例に限定されないことは言うまでもない。

本プログラムは、コネクト検索を行なうために必要となる分子名、サブネット名又は生体イベント名等の名前を得るための検索を行なう1101から1103までのステップと、コネクト検索を行ない、分子機能ネットワークを表示する1104から1108のステップ、及び生成した分子機能ネットワークにさらに処理を行なう付加的なステップ1109及び1110からなる。

利用者はまず、ステップ1101で分子名、分子略号、サブネット名、生体イベント名、病態イベント名、疾患名、アミノ酸配列、核酸塩基配列、外部データ

ベースID、医薬分子構造などに関する検索方法を指定し、質問文字列を入力する。検索方法としては、上記の項目に対して個別に検索を行なう方法や、複数の項目に対して共通の質問文字列で検索を行なう方法などから、利用者が選ぶことができる。質問文字列は、データベース中のデータ項目に完全に一致する文字列でなくてもよく、名前の一部分を表わす文字列やいわゆるワイルドカード文字などを含むものであってもよい。検索項目として蛋白質のアミノ酸配列又は核酸塩基配列を指定した場合には、利用者は質問文字列としてアミノ酸配列又は塩基配列を一文字コード（例：アラニン=A、グリシン=G、グアニン=g、シトシン=cなど）で表わした文字列を入力する。検索項目として医薬分子構造を指定した場合には、利用者はMOLFILEなどの形式で質問分子構造を表現したデータを入力する。

利用者が入力した検索項目に対して、プログラムはステップ1102で生体分子情報データベース、生体分子連鎖データベース及び関連するデータベースのデータ項目に対してキーワード検索、分子構造検索、配列検索などの方法により検索を行なう。キーワード検索では、文字列の完全な一致のみならず、文字列の部分的な一致やワイルドカードによる複数文字列への一致を許容してもよい。ステップ1101で検索項目としてアミノ酸配列又は塩基配列が指定された場合には、プログラムは生体分子情報データベース又は関連する配列データベース中のアミノ酸配列・塩基配列に対して、質問文字列（配列）の一致度や相同性による検索を行ない、一致度や相同性が高い配列のID又は対応する分子名を検索結果として与える。質問項目として医薬分子構造が指定された場合には、プログラムは部分構造マッチングの方法により部分構造が一致又は類似する医薬分子を検索し、該当する医薬分子名を検索結果として与える。

ステップ1102の検索で得られたヒット項目は、ステップ1103でリストとして表示される。プログラムは、リスト表示中の各ヒット項目を、それが分子名、サブネット名、生体イベント名のいずれに該当するかを、リスト中の表示位置を区分する方法や、アイコンを付加する方法などにより区別して表示する。

次に、利用者はステップ1104でコネクト検索の方法と検索の端点となる分子名、サブネット名又は生体イベント名（病態イベントも含む）を指定する。本実施例では、コネクト検索の方法として、1点を指定してその周辺につながるネットワークを検索する方法と、2点を指定してその間をつなぐネットワークを検索する方法が提供される。これらの2通りの検索方法で必要となる入力項目を、それぞれ第12図、第13図に示す。利用者は、ステップ1103で表示されたリストから適切な項目を選択することにより、分子名、サブネット名又は生体イベント名を入力する。該リスト中に適切な項目が存在しない場合には、利用者はステップ1101の検索項目の入力に戻って、適切な項目が見つかるまでステップ1101からステップ1103の検索工程を繰り返すことができる。

ステップ1105で、利用者はコネクト検索の制限条件を入力する。制限条件としては、生成される分子機能ネットワーク中に含まれる分子数の上限や、2点間の検索の場合には該2点間に介在する関連付けの数（パス数）の上限などが指定可能である。ステップ1106で、利用者は検索の結果得られる分子機能ネットワークの表示方法を指定する。表示方法としては、例えばネットワークを構成する全分子を明示的に表示する方法（分子ネットワーク表示）や、サブネットに属する分子をまとめて一つの節点として表示する方法（サブネット表示）、などの複数の方法から利用者が選ぶことができる。

ステップ1104からステップ1105で指定された条件に従って、プログラムはステップ1107で生体分子連鎖データベースに対してコネクト検索を行なう。検索の結果得られる分子機能ネットワークは、利用者がステップ1106で与えた表示方法の指定に従って、ステップ1108により分子、サブネット又は生体イベントを節点として持つグラフとして表示される。

利用者は、ステップ1108で表示された分子機能ネットワークを視覚的に検討し、必要に応じてステップ1104に戻ってコネクト検索の条件を変更して検索を繰り返したり、ステップ1101に戻って分子名、サブネット名又は生体イベント名の検索を繰り返したりすることもできる。

さらに、本プログラムでは付加的なステップ1109又はステップ1110により、生成された分子機能ネットワークに対してさらに処理を行なうことができる。ステップ1109では、利用者は複数の分子機能ネットワーク間の論理演算を行なうことができる。ステップ1109を実行するためには、ステップ1108までの工程を複数回実行して複数の分子機能ネットワークを生成しておく必要がある。これらの複数の分子機能ネットワークに対して、プログラムはネットワーク間で共通な部分（AND演算）や共通でない部分（XOR演算）を求めたり、複数のネットワークの和集合（OR演算）を求めたりすることができる。この機能は、異なる生物種や臓器などの間での分子機能ネットワークの違いを調べる目的などに有用である。

ステップ1110では、利用者は生成した分子機能ネットワークについて、さらに絞り込み検索を行なって、該分子機能ネットワーク中の分子や部分ネットワークを強調表示したり抽出表示したりすることができる。この絞り込み検索では、ステップ1101～1103で用いる検索方法のいずれも利用することができる。ステップ1110により、例えば特定の臓器で発現する生体分子を分子機能ネットワーク中で強調表示したり、広範囲な分子機能ネットワークから指定したサブネットに属するものだけを抽出表示したりすることが可能となる。

産業上の利用可能性

生体イベントを含む生体分子対の情報の集合である本発明の生体分子連鎖データベースは、検索によって必要な範囲の機能上生合成上の該分子間のつながりである分子機能ネットワークを生成し、任意の生体分子が発現に直接又は間接に関わる生体イベントを推定するのに有用であり、さらに医薬分子の情報又は遺伝情報とリンクすることにより、新薬開発や個体差に基づく医療に必要な知識を得ることができる。

請 求 の 範 囲

1. 生体イベントの情報を含む生体分子連鎖データベースを用いて、コネクト検索を行なうことにより、生体イベントを含む分子機能ネットワークを生成する方法。
2. 生体イベントの情報を含む生体分子連鎖データベースを用いて、コネクト検索を行なうことにより、生体イベントを含む分子機能ネットワークを生成し、該ネットワーク中の任意の生体分子と任意の生体イベントとの間の経路を推定する方法。
3. 生体イベントの情報を含む生体分子連鎖データベースを用いて、コネクト検索を行なうことにより、生体イベントを含む分子機能ネットワークを生成し、該ネットワーク中の任意の生体分子が関わる生体イベントを推定する方法。
4. 生体分子対の情報を階層化して保存した生体分子連鎖データベースを用いて、コネクト検索により分子機能ネットワークを生成する方法。
5. 請求の範囲第1項ないし第3項のいずれか1項に記載の方法であって、生体分子対の情報を階層化して生体分子連鎖データベースに保存することを特徴とする方法。
6. 生体分子及び／又は生体イベントの情報が階層化して保存されたデータベースを用いることを特徴とする、請求の範囲第1項ないし第5項のいずれか1項に記載の方法。
7. データベース中の項目に対するキーワード検索、分子構造検索、又は配列相同性検索のいずれか1つ又は2以上を併せて行なうことを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。
8. 請求の範囲第1項ないし第6項のいずれか1項に記載の方法であって、データベース中の項目に対してキーワード検索、分子構造検索、又は配列相同性検索のいずれか1つ又は2以上を行なうことによりコネクト検索に用いるデータを絞り込み、限定的な分子機能ネットワークを生成することを特徴とする方法。

9. 請求の範囲第1項ないし第6項のいずれか1項に記載の方法であって、生成した分子機能ネットワークに対して、さらにキーワード検索、分子構造検索、配列相同性検索のいずれか1つ又は2以上を行なうことにより、該ネットワークの部分ネットワークを生成する方法。

10. 生体イベントの情報にキイ分子の量的又は質的な変動に対応したアップオアダウンの情報を含むことを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

11. 生体イベントに疾患に関する病名、病態、診断基準、及び治療薬からなる情報のいずれか1つ又は2以上を含むことを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

12. 特定の生体分子を作用の標的とする医薬分子の情報を併せて用いることを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

13. 生体分子と遺伝子との対応関係の情報を併せて用いることを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

14. 臓器毎の蛋白発現情報及び／又は遺伝子発現情報を併せて用いることを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

15. 生体分子が遺伝子多型に関わる遺伝子情報とリンクされていることを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

16. 特定のキイ分子で発現が制御される遺伝子又は蛋白質の情報を併せて用いることを特徴とする、請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

17. 請求の範囲第12項の方法を用いることを特徴とする、医薬分子の副作用の推定方法。

18. 請求の範囲第1項ないし第16項のいずれか1項に記載の方法を用いることを特徴とする、創薬ターゲットの推定方法。

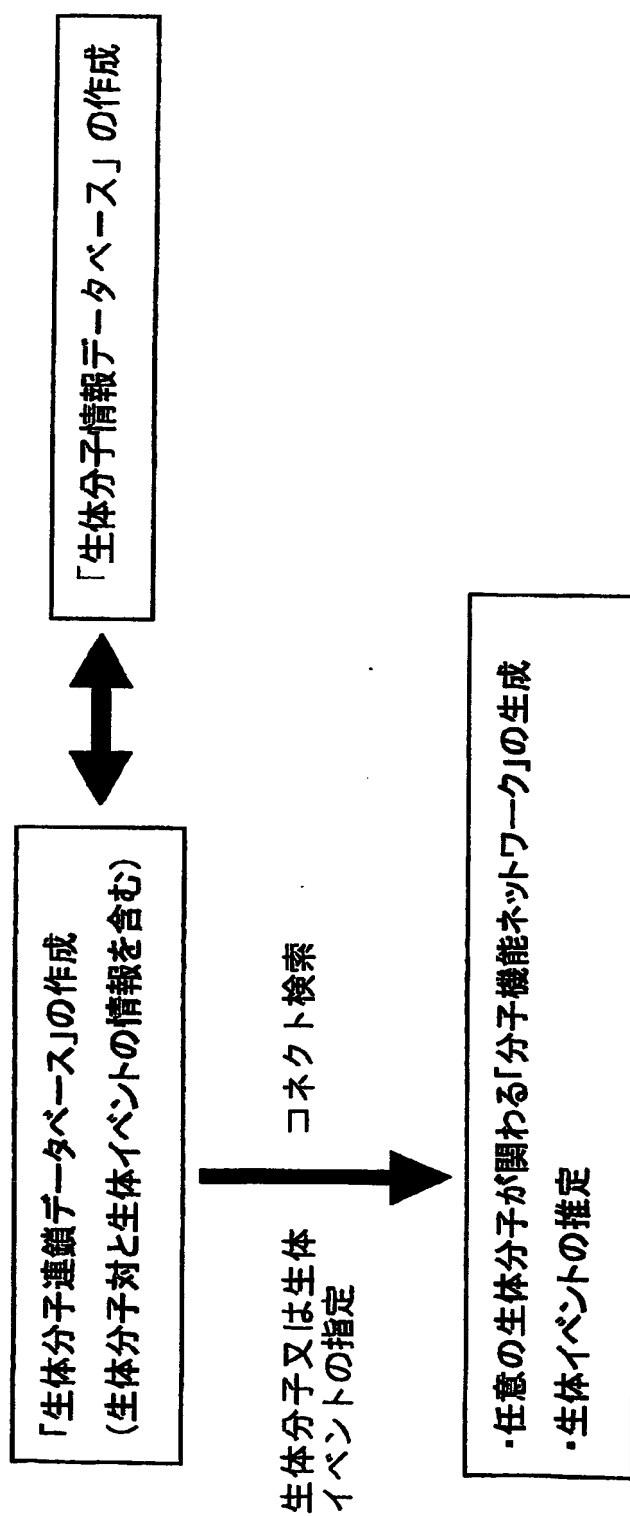
19. 請求の範囲第1項ないし第16項のいずれか1項に記載の方法を用いることを特徴とする、特定の生体分子を創薬ターゲットとした場合の副作用リスクの

推定方法。

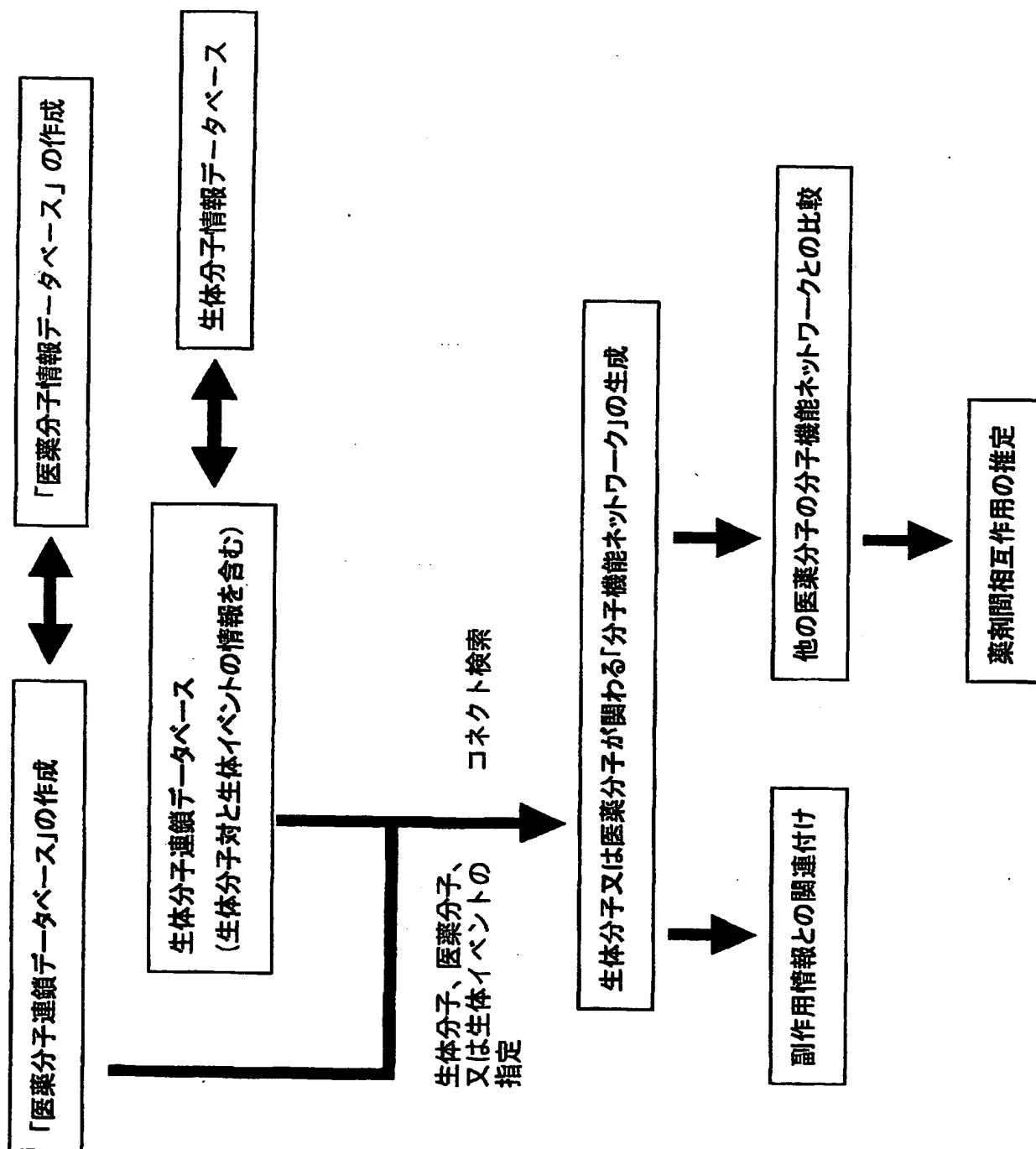
20. 請求の範囲第1項ないし第19項のいずれか1項に記載の方法を実施するためのプログラム及びデータベースからなるコンピュータシステム。

21. 請求の範囲第1項ないし第19項のいずれか1項に記載の方法を実施するためのプログラム及び／又はデータベースを記録したコンピュータ読み取り可能な媒体。

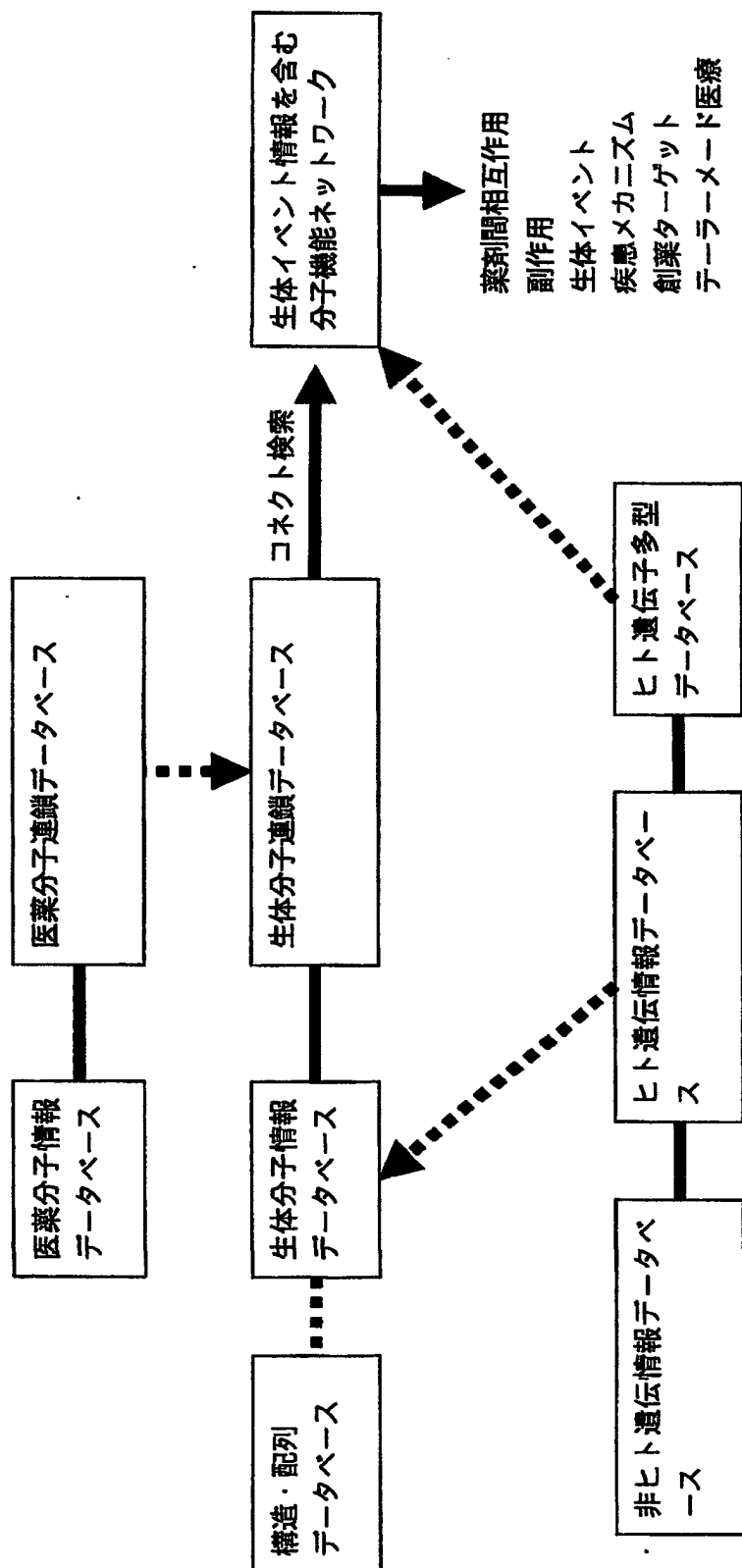
第1図



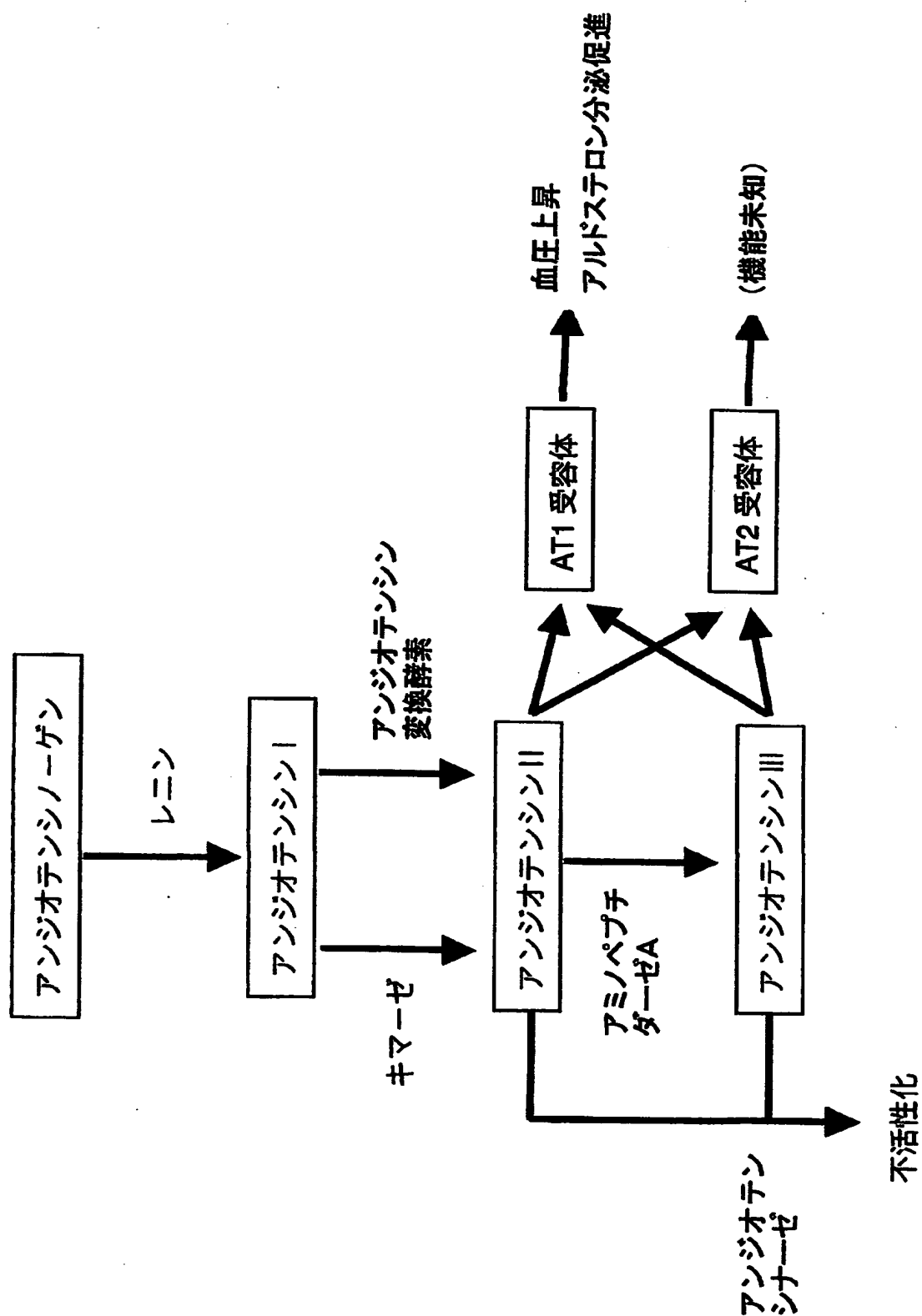
第2図



第 3 図



第4図



第5図

分子名	略号	分子コード	機能コード	SwissProt コード	生成臓器	存在臓器
レニン	REN	蛋白質	酵素	REN1_HUMAN	腎(傍糸球体細胞)	血中
アンジオテンシン変換酵素	ACE	蛋白質	酵素	ACE_HUMAN	血管内皮	血管内皮、血中
アンジオテンシナーゼ	PPCP	蛋白質	酵素	PCP_HUMAN	(リンゾーム)	(リンゾーム)
キマーゼ	CMA1	蛋白質	酵素	MCT1_HUMAN	マスト細胞、心臓	血中
アミノペプチダーゼ A	APA	蛋白質	酵素	AMPE_HUMAN	腎、ほか	腎、ほか
アンジオテンシノーゲン	ANGTS0	ペプチド	前駆体	ANGT_HUMAN	肝	血中
アンジオテンシン I	ANGTS1	ペプチド	前駆体	—	血中	血中
アンジオテンシン II	ANGTS2	ペプチド	活性ペプチド	—	血中	血中
アンジオテンシン III	ANGTS3	ペプチド	活性ペプチド	—	血中	血中
アンジオテンシン II 代謝物	ANGTX2	ペプチド	代謝産物	—	血中	—
アンジオテンシン III 代謝物	ANGTX3	ペプチド	代謝産物	—	血中	—
アンジオテンシン受容体 I	AGTR1	蛋白質	膜受容体	AG2R_HUMAN	血管、副腎	血管、副腎
アンジオテンシン受容体 II	AGTR2	蛋白質	膜受容体	AG22_HUMAN	脳、副腎	脳、副腎

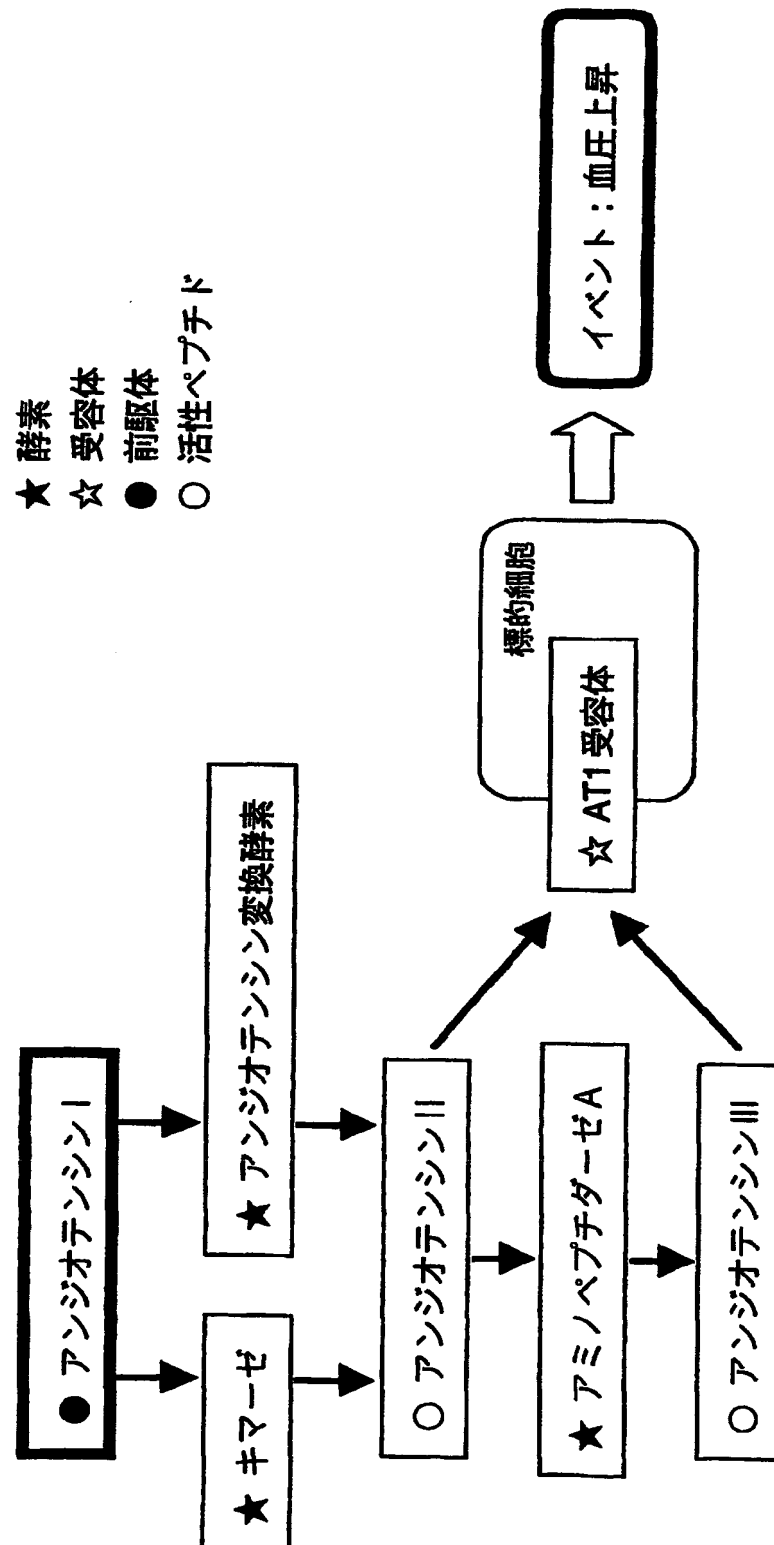
第6図

分子1	分子2	関係コード	作用臓器	生体イベント	病態イベント
ANGTNS0	REN	21	血中	-	-
REN	ANGTNS1	22	血中	-	-
ANGTNS1	ACE	21	血中(肺循環)	-	-
ACE	ANGTNS2	22	血中(肺循環)	-	-
ANGTNS1	CMA1	21	血中	-	-
CMA1	ANGTNS2	22	血中	-	-
ANGTNS2	APA	21	血中	-	-
APA	ANGTNS3	22	血中	-	-
ANGTNS2	PRCP	21	血中	-	-
PRCP	ANGTX2	22	血中	血圧降下	-
ANGTNS3	PRCP	21	血中	-	-
PRCP	ANGTX3	22	血中	血圧降下	-
ANGTNS2	AGTR1	11	血管、副腎	血管収縮、 血圧上昇	高血圧
ANGTNS2	AGTR2	11	脳、副腎	(不明)	(不明)

関係コード 11:アゴニスト-受容体、12:アンタゴニスト-受容体

21:基質-酵素、22:酵素-生成物、23:阻害剤-酵素、50:代謝産物

第7図



第8図

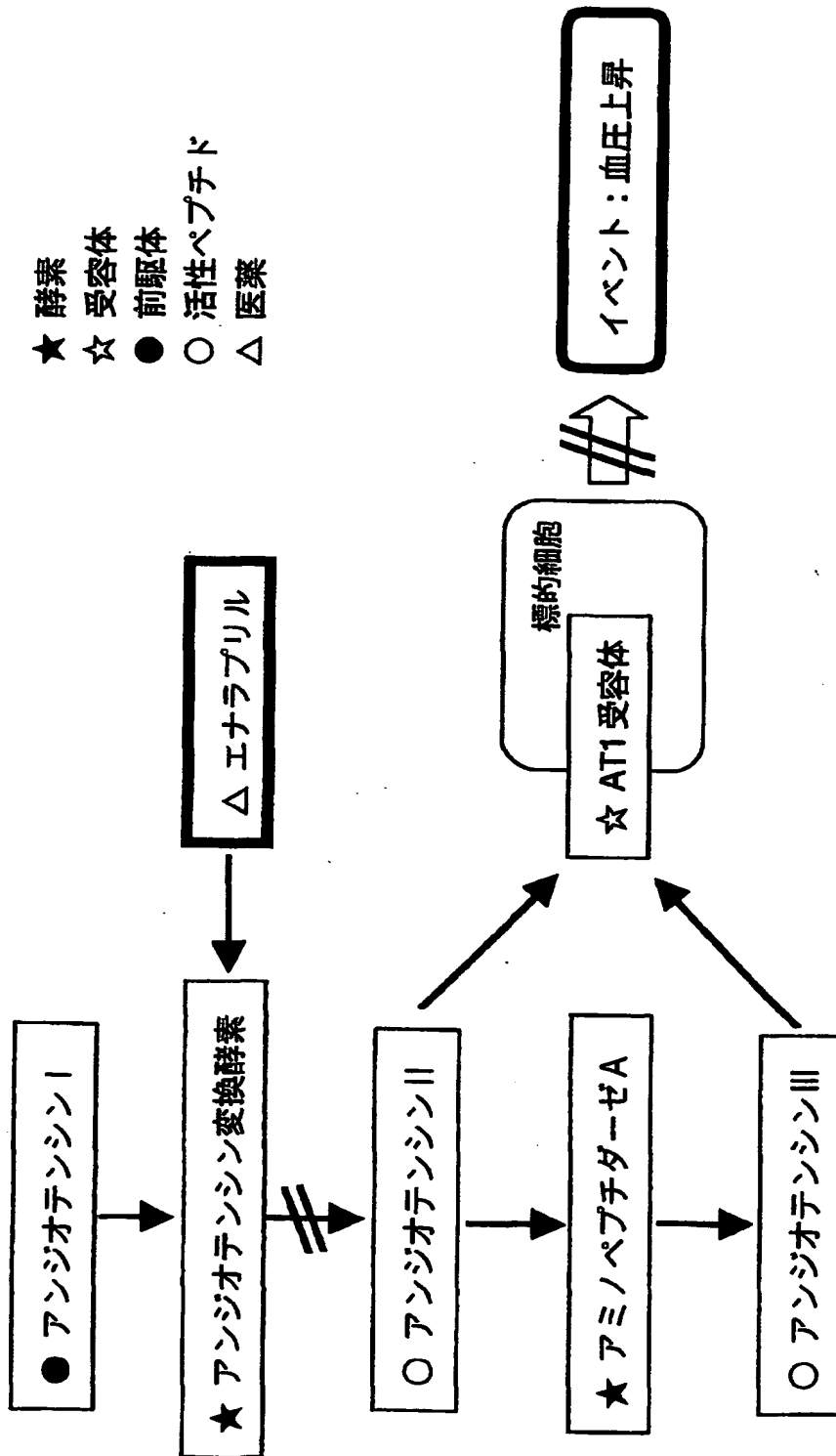
分子名	略号	作用	対象疾患	副作用
エナラプリル	ENALAPR	血圧降下	高血圧症、慢性心不全	空咳、急性腎不全、血管浮腫
ロサルタン	ROSARTN	血圧降下	高血圧症	血管浮腫、急性肝炎、腎不全

第 9 図

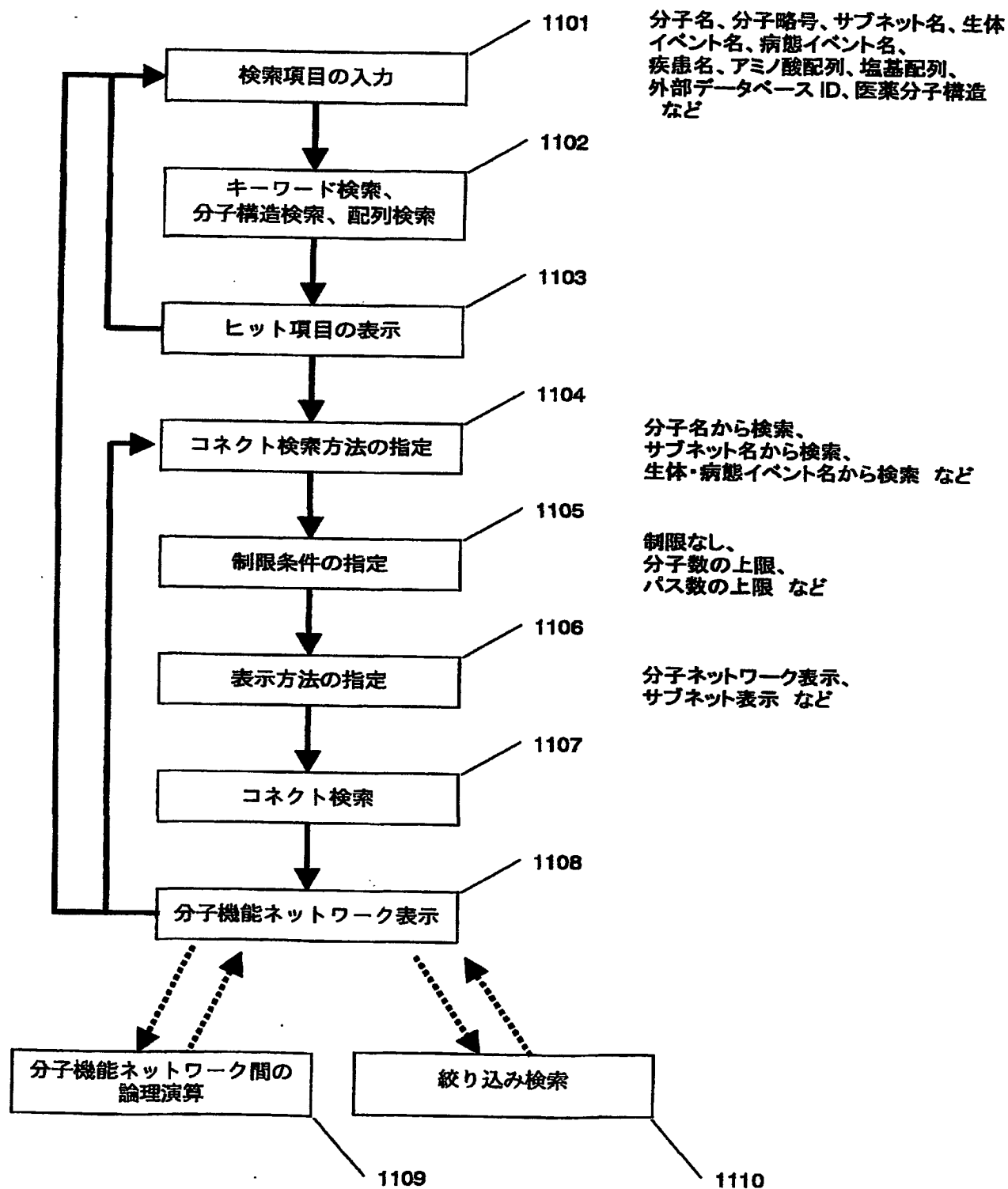
分子 1	分子 2	関係コード	作用臓器	薬理作用
ENALAPR	ACE	2 3	血中	血圧降下
ROSARTN	AGTR1	1 2	血管、副腎	血圧降下

関係コードの凡例 1 1 : アングiotensin II 受容体、1 2 : アングiotensin II 受容体
2 1 : 基質-酵素、2 2 : 酵素-生成物、2 3 : 阻害剤-酵素、5 0 : 代謝産物

第 10 図



第 1 1 図



第 1 2 図

入力項目(1点指定)
分子名
サブネット名
生体イベント名

第 1 3 図

入力項目(2点指定)	
分子名	分子名
分子名	サブネット名
分子名	生体イベント名
サブネット名	サブネット名
サブネット名	分子名
サブネット名	生体イベント名
生体イベント名	生体イベント名
生体イベント名	分子名
生体イベント名	サブネット名

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP01/07830

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl.⁷ G06F17/30, G06F17/60

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl.⁷ G06F17/30, G06F17/60

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2001
Kokai Jitsuyo Shinan Koho 1971-2001 Jitsuyo Shinan Toroku Koho 1996-2001

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST FILE (JOIS), WPI, INSPEC (DIALOG)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHIBAGAKI, Y. et al., "A graphical representation system assisting at gene regulatory network construction", Proceedings of the 14 th Annual Conference of JSAI, (2000), 03 July, 2000, pages 543 to 544	1-3, 13, 20, 21
Y	GOTO, S. et al., "Organizing and Computing metabolic pathway data in terms of binary relations", Pacific Symposium on Biocomputing, (2nd), Maui, (1996), pages 175 to 186, especially, page 179	4-12, 14-19
Y	HIRAYAMA, "Bunshi Sekkei no tameno Chishiki Data System no Koujika-hou no Kaihatsu; Lead Kagoubutsu Hakken no tameno Keiken-soku no Riron-ka ni kansuru Kenkyuu", Seitai Seigyo Busshitsu no Bunshi Sekkei to Seimitsu Gousei no tameno Kiban Gijutsu no Kaihatsu ni kansuru Kenkyuu (1 st period) Seika Houkokusho, Heisei 5-7 nendo (1993-1995) ed., (1996), pages 210 to 221	4-6
Y		7-12, 17-19

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
04 October, 2001 (04.10.01)

Date of mailing of the international search report
16 October, 2001 (16.10.01)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP01/07830

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	OKUBO et al., "Rensai Genome Data Base Keisan-ki ni Kakunou sareta Seimei no Sugata; Dai 6kai, Hatsugen Joho Database to Clustering; Idenshi no Furumai no Hyougen to sore ni motozuku Idenshi no Bunrui", bit, 01 December, 1999, Vol.399, pages 80 to 86	14-16
Y	NAKAI, "Genome Kinou Kenkyuu Protocol Micro Array, PCR, Bio-informatics no Saishin Gijutsu kara SNP, Model Seibutsu no Kaiseki made; Chapter 3, Hito Genome Bio-informatics; 1. Genome Informatics", Jikken Igaku separate volume, Post Genome Jidai no Jikken Kouza 1, 10 April, 2000, pages 163 to 167	14-16
A	KANEHISA, M., "Pathway Database and Logical Simulation of Genes and Molecules", Genome Science: Hito Genome Kaiseki ni motozuku Bio-Science no Shin-Tenkai, Heisei 8 nendo (1996), No.08283101, March, 1997, pages 218 to 221	1-21
A	KANEHISA, M., "Pathway Databases and Higher Order Function", Advances in Protein Chemistry, 06 June, 2000, (received at JICST), Vol.51, pages 381 to 408	1-21
PA	JP 2001-188768 A (Japan Science and Technology Corporation), 10 July, 2001 (10.07.01), & WO 01/48690 A1	1-21

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30, G06F17/60

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30, G06F17/60

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年

日本国公開実用新案公報 1971-2001年

日本国登録実用新案公報 1994-2001年

日本国実用新案登録公報 1996-2001年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)
WPI, INSPEC (DIALOG)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	柴垣 他「グラフィカル表現による遺伝子制御ネットワーク構築支援システム」人工知能学会全国大会(第14回)論文集, 3.7月. 2000, p. 543-544	1-3, 13, 20, 21
Y	SHIBAGAKI, Y. et al. A graphical representation system assisting at gene regulatory network construction. Proceedings of the 14th Annual Conference of JSAI, 2000	4-12, 14-19
Y	GOTO, S. et al. Organizing and Computing metabolic pathway data in terms of binary relations. Pacific Symposium on Biocomputing (2nd) Maui, 1996, p. 175-186, 特に, p. 179	4-6

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

04.10.01

国際調査報告の発送日

16.10.01

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

高瀬 勤



5M

9069

電話番号 03-3581-1101 内線 3599

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	平山「分子設計のための知識・データ・システムの高次化法の開発 リード化合物発見のための経験則の理論化に関する研究」生体制 御物質の分子設計と精密合成のための基盤技術の開発に関する研究 (第1期)成果報告書 平成5-7年度, 1996, p. 210-221	7-12, 17-19
Y	大久保, 他「連載 ゲノムデータベース-計算機に格納された生命の 姿 第6回 発現情報データベースとクラスタリング 遺伝子の振る 舞いの表現とそれに基づく遺伝子の分類」b i t, 1. 12月. 1999, Vol. 399, p. 80-86	14-16
Y	中井「ゲノム機能研究プロトコール マイクロアレイ, PCR, パイ オインフォマティクスの最新技術からSNP, モデル生物の解析ま で 3章ヒトゲノム III. パイオインフォマティクス 1. ゲノムイン フォマティクス」実験医学別冊 ポストゲノム時代の実験講座1, 10. 4月. 2000, p. 163-167	14-16
A	金久実「遺伝子・分子パスウェイデータベースと論理シミュレーシ ョン」ゲノムサイエンス: ヒトゲノム解析に基づくバイオサイエン スの新展開, 平成8年度, No. 08283101, 3月. 1997, p. 218-221 KANEHISA, M. Pathway Database and Logical Simulation of Genes and Molecules.	1-21
A	KANEHISA, M. PATHWAY DATABASES AND HIGHER ORDER FUNCTION. Advances in PROTEIN CHEMISTRY, 6. 6月. 2000 (JICST受入), Vol. 51, p. 381-408	1-21
P A	JP 2001-188768 A (科学技術振興事業団) 10. 7月. 2001 (10. 07. 01) & WO 01/48690 A1	1-21

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)